
Public Use Data File Documentation

2011-2013

National Survey of Family Growth

USER'S GUIDE

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Hyattsville, Maryland
December 2014

Data User's Agreement

Federal law (The Public Health Service Act (Section 308 (d))) provides that these data may be used only for the purpose of health statistical reporting and analysis. Any effort to determine the identity of any person or establishment is prohibited by this law.

NCHS does all it can to assure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted from the data files. In addition, a small number of records have had one or more responses slightly modified through statistical perturbation. These modifications do not affect point estimates or statistical significance, but are intended to prevent definitive identification of individual respondents.

By using National Survey of Family Growth data, you signify your agreement to comply with the following legal requirements:

1. To use these data for statistical reporting and analysis only;
2. To make no use of the identity of any person or establishment discovered inadvertently and advise the Director, NCHS, of any such discovery (301-458-4500); and
3. To not link these data with individually identifiable data from any other data set.

TABLE OF CONTENTS

<u>What You Need to Know about this User’s Guide</u>	5
<u>PART 1:</u>	
<u>GENERAL INFORMATION FOR USERS OF THE</u>	
<u>2011-2013 NSFG PUBLIC USE DATA</u>	7
<u>Background and Overview of the 2011-2013 NSFG</u>	8
<u>Where to Find NSFG Public Use Data and Documentation and Other NSFG Data Files</u>	9
<u>Organization of the 2011-2013 NSFG Public Use Data Files</u>	10
<u>File Characteristics</u>	10
<u>Data Layout for Each File</u>	11
<u>Sample Weights and Variance Estimation</u>	13
<u>Overview of Data Quality in the NSFG</u>	15
<u>Data Preparation for Public Use</u>	15
<u>Logical Inconsistencies and Out-of-Range Values</u>	15
<u>Coding for “Don’t Know,” “Refused,” and “Not Ascertained” Values</u>	16
<u>Century Month Coding for Dates</u>	17
<u>Other-Specify Coding</u>	21
<u>Recodes and Imputation</u>	21
<u>Protections to Minimize Risk of Disclosure of Individual-Level Data</u>	23
<u>Description of Codebooks</u>	25
<u>Overview</u>	25
<u>Elements of the Codebooks</u>	27
<u>Variable Name</u>	27
<u>Variable Type</u>	27
<u>Column Location</u>	28
<u>Question Text</u>	28
<u>Universe Statements (“Applicable specifications”)</u>	28
<u>Response Categories and Unweighted Frequencies</u>	29
<u>Link to Recode Specifications</u>	30
<u>Notes</u>	30
<u>Description of Questionnaires</u>	30
<u>CAPI-Lite Format</u>	30
<u>CAPI Reference Questionnaire (CRQ) Format</u>	30
<u>User Support</u>	33
<u>Acknowledgments</u>	33

<u>PART 2: TOPIC-SPECIFIC NOTES FOR ANALYSIS</u>	35
<u>Abortion Under-Reporting among Females in the NSFG</u>	35
<u>Birth Reporting among Females in the NSFG</u>	36
<u>Religion Data in the NSFG</u>	37
<u>Female Respondent File Notes</u>	40
<u>Female Pregnancy File Notes</u>	47
<u>Male File Notes</u>	49

APPENDICES (provided in separate PDF files posted on NSFG webpage)

Appendix 1: File Indexes for 2011-2013 NSFG Public Use Files

- 1a. Female Respondent File Index
- 1b. Female Pregnancy (Interval) File Index
- 1c. Male File Index

Appendix 2: SAS and STATA Syntax Guidelines for Common File Manipulations

- Combining data from female respondent and female pregnancy files
- Combining data from male and female files from 2011-2013
- Combining data across years of data collection
 - 1995, 2002, 2006-2010, and 2011-2013 for females
 - 2002, 2006-2010, and 2011-2013 for males

Appendix 3: Recode Specifications for 2011-2013 NSFG

- 3a. Female Respondent File Recode Specifications
- 3b. Female Pregnancy (Interval) File Recode Specifications
- 3c. Male File Recode Specifications

Appendix 4: Recode “Cross-walk” Grids

- 4a. Male-Female Recode Crosswalk for 2011-2013
- 4b. Female and Pregnancy Recode Crosswalks for 2002, 2006-2010, and 2011-2013
- 4c. Male Recode Crosswalk for 2002, 2006-2010, and 2011-2013

Appendix 5: Summary of NSFG questionnaire changes since 2006-2010

Appendix 6: Frequently Asked Questions about the NSFG

Appendix 7: List of Restricted Use Variables Available through the RDC

- 7a. List of Restricted-Use Analytic Variables for the Female File
- 7b. List of Restricted-Use Analytic Variables for the Male File
- 7c. Variables Modified or Suppressed for the Public Use Files

What You Need to Know about This User's Guide for the 2011-2013 NSFG

This User's Guide is for researchers planning to use the 2011-2013 National Survey of Family Growth (NSFG) 2011-2013 Public Use Data Files. It is intended to guide you through the process of identifying the variables you need for your research. The User's Guide consists of:

- **Part 1** providing an overview with information that applies to all users, regardless of their research topics.
- **Part 2** contains information about specific topics relevant to your analysis
- **7 Appendices** providing further reference material:
 - Appendix 1 – Indexes listing all variables on the public use data files
 - Appendix 2 – SAS and Stata syntax for common file manipulations
 - Appendix 3 – Specifications for all the recode (constructed) variables
 - Appendix 4 – Crosswalk grids for all recode variables and their equivalents by gender and across recent NSFG survey years
 - Appendix 5 – Summary of questionnaire changes since the 2006-2010 NSFG
 - Appendix 6 – Frequently Asked Questions (FAQ) for the NSFG
 - Appendix 7 – List of variables suppressed or modified for the public-use file including detail about restricted-use variables only available through the NCHS Research Data Center

As noted above, **Appendix 6** of this User's Guide provides a list of **frequently asked questions (FAQ)** about the NSFG, geared toward current or prospective users of the data file. Here are the primary highlights:

1. **Use sample weight (e.g., WGT2011_2013) and design variables (SECU & SEST)** to make valid estimates. Failure to use the weights and design variables correctly will lead to inaccurate findings.
2. **Use recoded variables when available.** They have been edited carefully, and missing data have been imputed. Some of the most commonly used recodes are listed on page 22.
3. In addition to this User's Guide with its 7 appendices, the **NSFG website** also provides:
 - a. the **questionnaires** in 2 formats
 - b. the **codebooks** (online Webdoc)
 - c. downloadable **ASCII data files**
 - d. **program statements** to read in the ASCII data using SAS, Stata, and SPSS
 - e. examples of variance estimation based on 2002 (Cycle 6) NSFG that can be adapted for 2011-2013
4. If you cannot find an answer to your question in the NSFG User's Guide, codebooks, questionnaire documents, or webpage, contact NSFG staff at nsfg@cdc.gov or 301-458-4222.
5. Until further information is published specific to 2011-2013 NSFG, refer to the section below on "Sample Weights and Variance Estimation," as well as the Series 1 and Series 2 technical reports from the 2006-2010 data release available on the NSFG website. The 2011-2013 NSFG sample design, fieldwork plan, and other procedures of the survey are

similar to what was used in 2006-2010.

For published reports or statistics based on NSFG data, please see the following:

- The NSFG website (www.cdc.gov/nchs/nsfg.htm) includes lists of **reports and articles using all Cycles of the NSFG**. NSFG reports published by the National Center for Health Statistics (NCHS) are accessible for download in PDF format from the NSFG website.
- **Another source** of published NSFG data is “**Key Statistics**,” accessible on the NSFG web page (http://www.cdc.gov/nchs/nsfg/key_statistics.htm). **Key Statistics** briefly present one or more of the most important statistics from nearly all topics covered in the NSFG. These statistics were originally posted on the webpage based on the 2002 NSFG and were then updated with the 2006-2010 NSFG. Within a few months of this public use file release, and also as topic-specific reports are published, the Key Statistics page will be updated with 2011-2013 NSFG data.

USER'S GUIDE PART 1: GENERAL INFORMATION FOR USERS OF THE 2011-2013 NSFG PUBLIC USE DATA

As noted above, the User's Guide for the 2011-2013 NSFG is divided into 2 parts.

Part 1 provides general information for users of the public use data, including:

- Overview of the 2011-2013 NSFG and brief background on previous NSFG surveys
- How to access NSFG data and documentation files
- How the NSFG data files are organized
- Information on sampling weights and variance estimation
- How the NSFG interview data were prepared for public use
- Descriptions of the codebooks and questionnaires

Part 2 ("**topic-specific notes for analysis**") gives information on data quality on selected areas and specific guidance for using particular variables in the data files.

The User's Guide also contains 7 appendices:

1. File indexes (i.e., listings of each variable with a brief description)
2. SAS & Stata syntax guidelines for common file manipulations
3. Recode specifications for specially constructed and imputed variables
4. Recode cross-walks by gender and across years of data collection
5. Summary of NSFG questionnaire changes from 2006-2010
6. Frequently Asked Questions about the NSFG
7. List of variables suppressed or modified for the public-use file including detail about restricted-use variables only available through the NCHS Research Data Center

BACKGROUND AND OVERVIEW OF THE 2011-2013 NSFG

The National Survey of Family Growth (NSFG) is designed and administered by the National Center for Health Statistics (NCHS), in collaboration with several other agencies of the US Department of Health and Human Services (see Acknowledgments). Since the NSFG began in 1973, there have been eight data file releases. The primary purpose of the survey is to produce national estimates of:

- factors affecting pregnancy, including sexual activity, contraceptive use, and infertility;
- the medical care associated with contraception, infertility, and childbirth;
- factors affecting marriage, divorce, cohabitation, and family building;
- adoption and caring for non-biological children;
- father involvement with their children;
- use of sexual and reproductive health services; and
- men's and women's attitudes about sex, childbearing, and marriage.

The survey results are used by the U.S. Department of Health and Human Services and other research and policy organizations to plan health services and health education programs, and to do statistical studies on the topics listed above, among others.

For the 2011-2013 NSFG, statistical design, interviewing, and data processing were conducted by the **University of Michigan's Institute for Social Research (ISR)**, under a contract with the National Center for Health Statistics (NCHS), in collaboration with the NCHS NSFG team. William Mosher served as Project Officer and Joyce Abma as Alternate Project Officer for the 2011-2013 data collection.

The following table presents basic information on each of the previous cycles of the NSFG, the 2006-2010 NSFG, and on the 2011-2013 data files, the focus of this User's Guide.

Cycle	Year	Scope	Number of Interviews	Over-Samples	Average Length	Incentive
1	1973	Ever-Married Women 15-44	9,797	Black Women	60 Minutes	No
2	1976	Ever-Married Women 15-44	8,611	Black Women	60 Minutes	No
3	1982	All Women 15-44	7,969	Black Women Teens	60 Minutes	No
4	1988	Women 15-44	8,450	Black Women	70 Minutes	No
5	1995	Women 15-44	10,847	Black Women Hispanic Women	100 Minutes	\$20
6	2002	Women 15-44 Men 15-44 (First time)	12,571 W = 7,643 M = 4,928	Blacks, Hispanics, 15-24 year olds	W= 85 min M= 60 min	\$40
n/a	2006-2010	Women 15-44 Men 15-44	22,682 W = 12,279 M = 10,403	Blacks, Hispanics, Teens	W=80 min M=60 min	\$40
n/a	2011-2013	Women 15-44 Men 15-44	10,416 W = 5,601 M =4,815	Blacks Hispanics Teens	W=80 min M=60 min	\$40

NSFG data are based on a multi-stage probability-based, nationally representative sample of the household population aged 15-44. Fieldwork for the 2011-2013 NSFG was conducted from September 2011 through September 2013. After sample respondents were selected based on screening interviews in selected households, in-person interviews were conducted with 5,601 women and 4,815 men 15-44 years of age for a total sample size of 10,416. The interviews were conducted by female interviewers trained specifically for the NSFG survey using laptop, computers---a procedure called computer-assisted personal interviewing (CAPI). The interviews for women averaged 80 minutes in length, and the interviews for men averaged 60 minutes, both within the limits approved by the Office of Management and Budget (OMB No. 0920-0314). The response rate for the 2011-2013 NSFG was 72.8% overall—73.4% for females, and 72.1% for males.

Additional details on how the survey was designed and conducted are forthcoming at the time of this writing. However the principles of continuous interviewing and many ongoing design features since it began in 2006 can be found in the following report:

RM Groves et al. Planning and Development of the Continuous National Survey of Family Growth. Vital and Health Statistics, Series 1(48). September 2009. Hyattsville, MD: National Center for Health Statistics. Available at http://www.cdc.gov/nchs/data/series/sr_01/sr01_048.pdf.

Also useful for details of design, sampling and operation, and statistical procedures, but more specific to the 2006-2010 data collection effort, are the following two reports:

JM Lepkowski et al. Continuous National Survey of Family Growth: Sample Design, Sampling Weights, Imputation, and Variance Estimation, 2006-2008. Vital and Health Statistics, Series 2(150). June 2010. Hyattsville, MD: National Center for Health Statistics. Available at: http://www.cdc.gov/nchs/data/series/sr_02/sr02_150.pdf

J. Lepkowski et al. Responsive Design, Weighting, and Variance Estimation in the 2006–2010 National Survey of Family Growth. Vital and Health Statistics, 2 (158). National Center for Health Statistics. June 2013. http://www.cdc.gov/nchs/data/series/sr_02/sr02_158.pdf

WHERE TO FIND NSFG PUBLIC USE DATA AND DOCUMENTATION and OTHER NSFG DATA FILES

The public use data and documentation for the 2011-2013 NSFG are available on the NSFG website (www.cdc.gov/nchs/nsfg.htm). Documentation includes this User's Guide, the questionnaires in two formats, and a link to the interactive, online codebook documentation ("Webdoc") for each variable on the files.

The public use data for the 2011-2013 NSFG are contained in three data files:

- **Female respondent** file (one record or observation per interviewed female)
- **Female pregnancy (interval)** file (one record per pregnancy of interviewed females)
- **Male respondent** file (one record per interviewed male)

Unlike in past NSFG file releases, when much of the data collected in the Audio Computer-Assisted Self Interviewing (ACASI) portion of the interview was released separately under a special data use agreement, the ACASI data are included in the main public use files for

2011-2013. The questionnaires available on the NSFG website provide more information on the specific content of the ACASI section. Essentially the items, now included in the main files, describe current and past behavior related to the risk of acquiring sexually transmitted diseases (STD) including the Human Immunodeficiency Virus, or HIV, the virus that causes AIDS.

For the 2006-2010 NSFG, the region of residence variable (REGION) was accessible only through the NCHS Research Data Center, due to concerns about confidentiality and the risk of disclosure if it were part of the public use files. For the 2011-2013 NSFG, sufficient measures were taken to modify or suppress other variables on the public-use files such that REGION can be made available upon request, using a special data user agreement similar to that used previously for the ACASI data files. Details for accessing REGION are available on the NCHS website.

Additionally, a 2011-2013 NSFG respondent-based file with the state and county FIPS codes is accessible through the NCHS Research Data Center to enable researchers to merge the NSFG survey data with other contextual data sources. An interviewer observation file is not provided for the 2011-2013 NSFG at this time, but will be compiled once four years of interviews have been collected, covering 2011-2015.

ORGANIZATION OF THE 2011-2013 NSFG PUBLIC-USE DATA FILES

The public-use data for the 2011-2013 NSFG are provided as 3 separate ASCII files. The file characteristics of these ASCII files are shown below.

FILE CHARACTERISTICS	Number of Records (observations)	Record Length (number of columns)	Number of Variables
Female respondent file File = 2011_2013_FemRespData.dat (one record per woman)	5,601	4,957	3,094
Female pregnancy (interval) file File = 2011_2013_FemPregData.dat (one record per pregnancy)	9,547	479	278
Male respondent file File = 2011_2013_MaleData.dat (one record per man)	4,815	4,636	3,059

The **Female Respondent file** contains one record for each of the 5,601 women in the survey and includes most of the information from their interviews. The **Female Pregnancy (Interval) file** contains one record for each of 9,547 pregnancies (both completed pregnancies and current pregnancies), and contains information about the characteristics of each pregnancy and method use and wantedness before each pregnancy. That is, in the Female Respondent file the unit of analysis is the woman, and in the Pregnancy file the unit of analysis is the pregnancy or pregnancy interval. The third data file, the **Male Respondent** file, often just referred to as the male file, contains one record for each of the 4,815 men interviewed; in this file the male respondent is the unit of analysis.

These ASCII data files can be read into any statistical software system for analysis purposes. Program statements to read the data into SAS, SPSS, and Stata, including variable and value labels, are provided on the NSFG webpage. In addition, SAS and Stata syntax guidelines are provided in **Appendix 2** for common file manipulations such as combining data for males and females or combining female respondent-level and pregnancy-level data.

Data Layout for Each File

The following is a listing of the beginning column locations of the major sections and key variables contained in all 3 public use files from the 2011-2013 NSFG.

- For more detail on data file contents than what is shown below, see **Appendix 1** (File Indexes) in which short descriptions and variable types are provided for every variable included on the public use files.
- More detailed outlines of the male and female questionnaires, as well as the full questionnaires fielded in 2011-2013, are provided on the NSFG website.
- Though the male and female questionnaires were organized somewhat differently, they included much of the same information. **Appendix 4** provides a “crosswalk” of recodes for males and females, as well as crosswalks showing comparable recodes over NSFG public use files 2002 and later.

FEMALE RESPONDENT FILE – information for each female

Beginning

Column # Items or Sections

1	Respondent ID (CASEID) and selected screener variables
11	Questionnaire Data (including computed variables) for Sections A-J
11	A: Background and demographic information
103	B: Pregnancy and adoption-related information
654	C: Marital and relationship history; first sexual intercourse; recent partners
1561	D: Sterilizing operations and impaired fecundity
1756	E: Contraceptive history and related information
3099	F: Family planning and medical services
3282	G: Birth desires and intentions

3301	H: Infertility services and reproductive health
3464	I: More background, more demographic information, & attitude questions
3544	J: Audio CASI: general health measures; pregnancy reporting; drug use; STD/HIV-risk behaviors; non-voluntary intercourse; income and economic insecurity
3751	Recodes (constructed, imputed variables) & imputation flags (including key recodes describing pregnancies)
4906	Weights & related variables (<i>including sample design variables</i>)
4926	Date of interview and related variables

FEMALE PREGNANCY (INTERVAL) FILE – information for each pregnancy (including current)

Beginning Column #	Items or Sections
1	Respondent ID (CASEID)
6	Pregnancy Order (PREGORDR)
8	Questionnaire Data (including computed variables) for Sections B & E
8	B: pregnancy outcomes and dates, prenatal care, sources of payment for delivery, breast-feeding
244	E: contraceptive use in the pregnancy interval and wantedness of the pregnancy
313	Recodes (constructed, imputed variables) & imputation flag (including key recodes and other variables from respondent file)
434	Weights & related variables (<i>including sample design variables</i>)
454	Date of interview and related variables

MALE RESPONDENT FILE – Information for each male

Beginning Column #	Items or Sections
1	Respondent ID (CASEID) and selected screener variables
11	Questionnaire Data (including computed variables) for Sections A-K
11	A: Background and demographic information
113	B: Sex education, vasectomy & infertility, sexual intercourse, enumeration and relationship with up to 3 recent (or last) sexual partners
243	C: Current wife or cohabiting partner: date of marriage; children; contraception with her
857	D: Recent (or last) sexual partner(s) (up to three): key dates, children; contraception with her; 1 st partner
1842	E: Former wives and first premarital cohabiting partner: key dates, children; contraception with each
2983	F: Other biological and adopted children; other pregnancies
3667	G: Fathering: Activities with the youngest child he lives with and the youngest child he lives apart from

3750	H: Desires and intentions for future children
3768	I: Health conditions, access to health care, and receipt of health services
3913	J: More background, more demographic information, & attitude questions
3983	K: Audio CASI: pregnancy reporting; drug use; STD/HIV-risk behaviors; non-voluntary intercourse; income and economic insecurity
4227	Recodes (constructed, imputed variables) & imputation flags
4593	Weights & related variables (<i>including sample design variables</i>)
4613	Date of interview and related variables

SAMPLE WEIGHTS AND VARIANCE ESTIMATION

Since the NSFG data are based on a multi-stage probability-based, nationally representative sample of the household population aged 15-44 data users should understand how to account for the complex sample design when doing their analyses in order to obtain statistically valid results. This section provides a summary of the procedures used for sample weighting and variance estimation. Since the sample design for the 2011-2013 NSFG is the same as that for the 2006-2010 much of the information produced for that file release is applicable for 2011-2013. For example:

J Lepkowski et al. 2010. The 2006-2010 National Survey of Family Growth: Sample Design and Analysis of a Continuous Survey. *Vital and Health Statistics* 2(150). National Center for Health Statistics. June 2010. http://www.cdc.gov/nchs/data/series/sr_02/sr02_150.pdf

J. Lepkowski et al. 2013. Responsive Design, Weighting, and Variance Estimation in the 2006–2010 National Survey of Family Growth. *Vital and Health Statistics*, 2(158). National Center for Health Statistics. June 2013. http://www.cdc.gov/nchs/data/series/sr_02/sr02_158.pdf

At the time of this writing, updates of this information are being prepared for the 2011-2013 NSFG. You will find the most pertinent information for conducting your analyses.

Sampling Weights

In the NSFG, Hispanics, blacks, and teens were selected at higher rates than others in the 15-44 age group. Women also had a higher probability of selection than men. Each person in the sample represents a different number of people in the population. The number they represent is called a sampling weight. Sampling weights adjust for these unequal probabilities of selection as well as for differential response rates and coverage rates, so that accurate national estimates can be made from the sample. The weights were adjusted to U.S. Census Bureau projections of the number of persons in age-sex-race-ethnicity subgroups. We strongly recommend using the weights in all analyses. Using the weights will permit replication of the nationally representative estimates that appear in published NCHS reports.

The 2011-2013 data files each have one weight variable called “WGT2011_2013” applying to all 10,416 interviews (5,601 female interviews and 4,815 male interviews) conducted from September 2011 through September 2013. This “WGT2011_2013” variable yields estimates representative of the 61 million men and 61 million women in the household population aged 15-44 of the United States at the midpoint of 2011-2013 interviewing (July 2012).

To yield number in thousands, as often appears in NCHS reports, you divide the sample weight by 1,000. For example, you would use this data manipulation step to create a new weight variable:

$$\text{WGT1000} = \text{WGT2011_2013} / 1000$$

In addition to using the sampling weight variable WGT2011_2013, researchers must use the design variables for the sampling stratum (SEST) and cluster (SECU) to obtain correct standard errors for their estimates.

Variance Estimation

Sampling variance is a measure of the quality of a statistic (such as a percentage, proportion or a mean) caused by having taken a sample instead of interviewing the full population. In the 2011-2013 NSFG, the sampling variance measures variation caused by interviewing the NSFG sample instead of all women and men aged 15-44 in the entire country.

Many statistical software packages by default, compute “population” variances, which may severely underestimate or overestimate the sampling variances because they assume that the sample was drawn using simple random sampling. Special software is required to accurately estimate sampling errors in a complex sample such as the NSFG; such software is increasingly common and easy to use. For example, SAS procedures such as “FREQ” produce population variances assuming simple random sampling, but SAS has also developed procedures for complex survey estimates in its ‘SURVEY’ procedures such as “SURVEYFREQ.” Similarly, SUDAAN, Stata and SPSS have procedures designed to analyze data derived from a complex sample survey.

Selected examples of how to use such statistical software to estimate sampling errors based on 2011-2013 NSFG data are provided in **Appendix 2**. Variance estimation examples using Cycle 6 (2002) NSFG are posted on the NSFG web site http://www.cdc.gov/nchs/nsfg/nsfg_cycle6.htm, and can easily be adapted to use 2011-2013 data.

When estimating variances for population subgroups (such as those who have ever had sexual intercourse or those 30-44 years of age), it is important to read in the entire data set first. An indicator variable (e.g. in SAS for 30-44 year olds, “if ager >= 30 then agenew=1; else agenew=2;”) should be created to identify only those observations that will be used in the analysis. Then, you define your subgroup of interest within your survey procedure, such as using the SUBPOPN command in SUDAAN, the SUBPOP command in STATA or the DOMAIN statements in SAS. If the data are subset without first reading in the entire data set, then empty clusters may be lost, making the survey design structure incomplete. This may result in the statistical software to terminate or lead to other errors.

OVERVIEW OF DATA QUALITY IN THE NSFG

As measured by amounts of missing data and inconsistent data, data quality in the 2011-2013 NSFG is high, as it was in previous data years. This high quality was obtained through:

- questionnaire design work, including careful specification, testing, and incorporation of lessons learned from the 2006-2010 NSFG;
- consistency checks built into the interview that allowed potential data problems to be resolved in the field rather than after data collection;
- evaluation of monthly data files to find and correct instrument problems before significant numbers of cases were affected; and
- extensive interviewer training to ensure adherence to consistent and ethical fieldwork procedures.

Among the thousands of variables in the file, Part 2 of the User's Guide provides information on those few variables where data issues exist.

Data files as large and complex as these cannot be guaranteed to be free of errors. If you believe you have found an error or need assistance apart from what is discussed in this User's Guide, please email the NSFG staff at NCHS at nsfg@cdc.gov.

DATA PREPARATION FOR PUBLIC USE

This section describes steps taken to prepare the NSFG interview data for public use. Some of these actions were taken simply to make the data more useful. Other actions were taken to protect the confidentiality of individual respondents, in keeping with the legal obligation of NCHS when conducting the NSFG or any of its other surveys.

Logical Inconsistencies and Out-of-Range Values

During fieldwork, logical consistency across data items was maintained through "edit checks" built into the programs that ran the male and female questionnaires. These edit checks alerted the interviewer to inconsistent or out-of-range entries and required that she attempt to correct the entry, usually by working with the respondent. Out-of-range values are greatly reduced in the 2011-2013 NSFG (as in 2006-2010) because valid ranges are specified and programmed into the instrument to the extent possible, and values outside that range are rejected or signaled by the computer.

Some edit checks in the instrument are "**hard edits**" in that they disallow combinations of values that are impossible (for example, respondents cannot report a date for any event in their lives that is later than the interview date or before their date of birth). Other edit checks are "**soft edits**" in that they alert the interviewer to situations that are rare but not impossible (for example, a respondent reports that she had her first menstrual period at a particularly young age).

In soft edit checks, the respondent is given the opportunity to revise his or her responses

in case they were given in error. If the respondent says that the information is accurate, the interviewer is allowed to override or suppress the inconsistency warning box and enter a brief comment to explain the situation. In rare cases, the interviewer herself may have misunderstood the edit check and mistakenly suppressed it. In all such cases, the seemingly inconsistent data may remain on the data file. It is not possible to foresee and specify all the edit checks that might be needed in these very complex interviews, and as a result, some inconsistencies in the data could not be eliminated.

In addition to edit checks, other aspects of the questionnaire designed to maximize consistency *during* data collection were: 1) “summary screens” before or after key sections, reminding the respondent of events and dates reported earlier, and 2) life-history calendars provided to female respondents as a visual aid for recording and remembering the chronology of events.

In the 2011-2013 NSFG, the process of checking for consistency by NCHS and ISR staff was focused primarily on the recoded variables and variables used to construct them. These were considered to be the most critical and most frequently used variables in the files. Considerable efforts were made to detect and resolve or document inconsistencies and unacceptable codes throughout the files. However, given the size and complexity of these data files, they may not be free of inconsistent or missing responses.

Coding for “Don’t Know,” “Refused,” and “Not Ascertained” Values

Missing data refers to responses of “don’t know” or “refused” that were entered by the interviewer to indicate that the respondent could not or would not provide an answer to a question. “Not ascertained” refers to rare instances in which a question was erroneously skipped during the interview. The code for “not ascertained” was generally assigned in these cases after fieldwork was completed. Only completed cases are in the files; a case was defined as being complete if the respondent answered the last applicable question in Section I (females) and Section J (males). The small number of respondents who did not complete the ACASI section, partially or completely, will have “not ascertained” values from their break-off point.

Depending on the column length of the original data items:

- “don’t know” values are coded 9, 99, 999, 9999, or 99999
- “refusal” values are coded 8, 98, 998, 9998, or 99998
- “not ascertained” values are coded 7, 97, 997, 9997, or 99997

(The codebooks only show these codes if the variable has cases with those particular values.)

Missing data as described above is distinct from a variable that was inapplicable -- the respondent was legitimately skipped past the question (for example, respondents who had never been pregnant were not asked questions about how their pregnancies ended). For more information on determining who was asked each question, refer to the description of universe statements in the User’s Guide section entitled “**Description of Codebooks**” further below or the codebook entry for particular variables. A question that was legitimately skipped or a variable legitimately not defined for a respondent will be coded as blank, and in the codebook, is

indicated by a “dot” and labeled “inapplicable” or “sysmis.”

Recoded variables may have legitimate inapplicable values, but they have no missing data in the form of “don’t know,” “refused,” or “not ascertained” values because these responses were imputed to a valid value. Cases that had recode values imputed because of missing information on the source variables are identified with an imputation “flag”-- a separate variable that indicates whether or not the corresponding recode was imputed (see User’s Guide section on “**Recodes and Imputation**” further below, as well as **Appendix 3** (Recode Specifications)).

Century Month Coding for Dates

During the interview, dates of events were collected as month and year, except for the respondent’s date of birth, which was collected as month, day, and year. For every date asked in the interview, including the respondent’s date of birth, the month and year information was converted to “century months” by subtracting 1900 from the year, then multiplying the remainder by 12, and adding the number of the month, where January = 1, February = 2, and so on. *(Note that the respondent’s day of birth is not available on the public use file. His or her birthday is converted to the century month variable cmbirth.)*

For instance:

The century month code for February 1959 is $(59 \times 12) + 2 = 710$.

The century month code for October 1987 is $(87 \times 12) + 10 = 1054$.

The century month code for January 2000 is $(100 \times 12) + 1 = 1201$.

The century month code for July 2006 is $(106 \times 12) + 7 = 1279$.

The century month form is convenient for computing intervals between dates, and subtraction yields intervals in months.

With the exception of one recoded date variable (DATEUSE1 on the female respondent file) that has a leading 9 to indicate when the value was estimated, all century month date variables in the file are 4 columns long. The following codes were used for the 3 types of missing data on century month date variables:

9997 = Not ascertained

9998 = Refused

9999 = Don’t know

When seasons are reported on any month variable in the NSFG interview, a particular month is consistently assigned to enable the construction of a century month value and facilitate subsequent routing through the questionnaire. Listed below are the months assigned when a season is reported:

Winter = 1 (January)

Spring = 4 (April)

Summer = 7 (July)
Fall = 10 (October)

If a respondent says “don’t know” or refuses (DK/RF) to report a month, the value “6” (June) is assigned for the month. If a respondent does not report a year, the century month variable is set to 9999 for “Don’t Know” or 9998 for “Refused.” Data users may wish to refer to the raw month and year variables included on the public use files, particularly the month variable, if they encounter an implausible or less likely set of century month dates.

The century month codes from 277 (January 1923) through 1368 (December 2013) are shown in the array below with the years from 1923 through 2013 on the vertical axis and the months on the horizontal axis. The code for a given month and year can be found by reading across the line for the appropriate year to the column headed by the appropriate month.

All interviews for the 2011-2013 NSFG were conducted between September 2011 (century month 1341) and September 2013 (century month 1365).

DATE CODES

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1923	277	278	279	280	281	282	283	284	285	286	287	288
1924	289	290	291	292	293	294	295	296	297	298	299	300
1925	301	302	303	304	305	306	307	308	309	310	311	312
1926	313	314	315	316	317	318	319	320	321	322	323	324
1927	325	326	327	328	329	330	331	332	333	334	335	336
1928	337	338	339	340	341	342	343	344	345	346	347	348
1929	349	350	351	352	353	354	355	356	357	358	359	360
1930	361	362	363	364	365	366	367	368	369	370	371	372
1931	373	374	375	376	377	378	379	380	381	382	383	384
1932	385	386	387	388	389	390	391	392	393	394	395	396
1933	397	398	399	400	401	402	403	404	405	406	407	408
1934	409	410	411	412	413	414	415	416	417	418	419	420
1935	421	422	423	424	425	426	427	428	429	430	431	432
1936	433	434	435	436	437	438	439	440	441	442	443	444
1937	445	446	447	448	449	450	451	452	453	454	455	456
1938	457	458	459	460	461	462	463	464	465	466	467	468
1939	469	470	471	472	473	474	475	476	477	478	479	480
1940	481	482	483	484	485	486	487	488	489	490	491	492
1941	493	494	495	496	497	498	499	500	501	502	503	504
1942	505	506	507	508	509	510	511	512	513	514	515	516
1943	517	518	519	520	521	522	523	524	525	526	527	528
1944	529	530	531	532	533	534	535	536	537	538	539	540
1945	541	542	543	544	545	546	547	548	549	550	551	552
1946	553	554	555	556	557	558	559	560	561	562	563	564
1947	565	566	567	568	569	570	571	572	573	574	575	576
1948	577	578	579	580	581	582	583	584	585	586	587	588
1949	589	590	591	592	593	594	595	596	597	598	599	600
1950	601	602	603	604	605	606	607	608	609	610	611	612
1951	613	614	615	616	617	618	619	620	621	622	623	624
1952	625	626	627	628	629	630	631	632	633	634	635	636
1953	637	638	639	640	641	642	643	644	645	646	647	648
1954	649	650	651	652	653	654	655	656	657	658	659	660
1955	661	662	663	664	665	666	667	668	669	670	671	672
1956	673	674	675	676	677	678	679	680	681	682	683	684
1957	685	686	687	688	689	690	691	692	693	694	695	696
1958	697	698	699	700	701	702	703	704	705	706	707	708
1959	709	710	711	712	713	714	715	716	717	718	719	720
1960	721	722	723	724	725	726	727	728	729	730	731	732
1961	733	734	735	736	737	738	739	740	741	742	743	744
1962	745	746	747	748	749	750	751	752	753	754	755	756
1963	757	758	759	760	761	762	763	764	765	766	767	768
1964	769	770	771	772	773	774	775	776	777	778	779	780
1965	781	782	783	784	785	786	787	788	789	790	791	792

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1966	793	794	795	796	797	798	799	800	801	802	803	804
1967	805	806	807	808	809	810	811	812	813	814	815	816
1968	817	818	819	820	821	822	823	824	825	826	827	828
1969	829	830	831	832	833	834	835	836	837	838	839	840
1970	841	842	843	844	845	846	847	848	849	850	851	852
1971	853	854	855	856	857	858	859	860	861	862	863	864
1972	865	866	867	868	869	870	871	872	873	874	875	876
1973	877	878	879	880	881	882	883	884	885	886	887	888
1974	889	890	891	892	893	894	895	896	897	898	899	900
1975	901	902	903	904	905	906	907	908	909	910	911	912
1976	913	914	915	916	917	918	919	920	921	922	923	924
1977	925	926	927	928	929	930	931	932	933	934	935	936
1978	937	938	939	940	941	942	943	944	945	946	947	948
1979	949	950	951	952	953	954	955	956	957	958	959	960
1980	961	962	963	964	965	966	967	968	969	970	971	972
1981	973	974	975	976	977	978	979	980	981	982	983	984
1982	985	986	987	988	989	990	991	992	993	994	995	996
1983	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008
1984	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020
1985	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032
1986	1033	1034	1035	1036	1037	1038	1039	1040	1041	1042	1043	1044
1987	1045	1046	1047	1048	1049	1050	1051	1052	1053	1054	1055	1056
1988	1057	1058	1059	1060	1061	1062	1063	1064	1065	1066	1067	1068
1989	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080
1990	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092
1991	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104
1992	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116
1993	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128
1994	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140
1995	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152
1996	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164
1997	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176
1998	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188
1999	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200
2000	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212
2001	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224
2002	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236
2003	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248
2004	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260
2005	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272
2006	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284
2007	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296
2008	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308
2009	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320
2010	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332
2011	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344
2012	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356
2013	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368

Other-Specify Coding

In the 2011-2013 NSFG, a small number of questions contained items to which respondents could specify a response other than those provided, and this response was typed in verbatim by the interviewer. In most cases, these questions appear in the CRQs with a “_SP” at the end of their variable name or a “SP_” at the beginning, and the question text and response type distinguish these clearly as open-ended questions. In all cases, responses that were clearly examples of a pre-existing category (part of the questionnaire instrument and shown to respondents on a show card) were recoded (or “back-coded”) into that pre-existing category. The remaining unique responses were examined by NCHS staff and if the numbers were sufficient, the new response categories were added to the list of existing (pre-specified) response categories for subsequent interviews. In a few instances, the responses were categorized and a new variable was created that contained only these additional responses.

Please consult the questionnaires for further details on these questions with “other (specify)” options. Part 2 of the User’s Guide also includes specific information related to some of these items. The actual verbatim responses are not included in the public use data files, but they have all been reflected in some way in the associated, closed-ended variables included on the public use file. Questions where new categories or new variables were created based on other-specify responses have a short description in a “Note” entry on their codebook page (see also **Description of Codebooks** further below).

Recodes and Imputation

*(also see **Appendix 1** (File Indexes), **Appendix 3** (Recode Specifications), and **Appendix 4** (Recode Crosswalks))*

In order to facilitate consistent, comparable estimates of key NSFG measures, NCHS produces a number of “recoded variables,” or “recodes” for each public use file. Published NCHS reports use these recodes whenever available because they permit internally consistent and replicable estimates. NCHS also uses the recodes to prioritize the cleaning of the data file: there are too many variables in the data file to edit or reconcile them all, so NCHS focuses its cleaning and editing primarily on the recodes and on the variables that are used to construct the recodes. (Recodes comprise about 10% of the variables in these files.)

Some recodes are fairly simple, while others are quite complex. Some recodes may simply be transferred from single questionnaire items and imputed if missing (for example, RCURPREG, whether respondent is currently pregnant). Other recodes are based on multiple questionnaire items and may involve more intricate logic to define. (For example, CONSTAT1, the respondent’s current contraceptive status)

Before using the original data items or constructing their own summary variables, analysts are encouraged to **check to see if a relevant recode exists**. Many of the raw or computed variables that have a recode corresponding to them will have a **note on their codebook pages** stating the name of the appropriate recode.

For convenience, below is a list of some commonly used recodes corresponding to background characteristics and other key NSFG variables. Unless otherwise indicated, the recodes are available for males and females.

AGER	R's age at interview
FMARITAL	Formal (legal) marital status
RMARITAL	Informal marital status
EDUCAT	Education (number of years of schooling)
HIEDUC	Highest completed year of school or highest degree received
HISPANIC	Hispanic origin, regardless of race
RACE	Race of respondent, regardless of Hispanic origin
HISPRACE	Race and Hispanic origin – based on 1977 OMB guidelines
HISPRACE2	Race and Hispanic origin – based on 1997 OMB guidelines
INTCTFAM	Intact status of childhood family
PARAGE14	Parental living situation at age 14
EDUCMOM	Mother's (or mother-figure's) education
AGEMOMB1	Age of mother (or mother-figure) at first birth
METRO	Place of residence (metropolitan-nonmetropolitan)
RELIGION	Current religious affiliation
LABORFOR	Labor force status
POVERTY	Poverty level income
TOTINCR	Total income of R's family
PUBASSIS	Whether R received public assistance in the calendar year before the interview
HADSEX	Whether R has ever had sexual intercourse with opposite sex
VRY1STAG	Age at first intercourse
VRY1STSX	Date (century month) of first intercourse
CONSTAT1	Current contraceptive status (females only)

Besides the above list, other sources to check are the File Indexes in **Appendix 1** or the Recode Specifications in **Appendix 3** to see if a relevant recode exists. There is also a section in the online codebooks (Webdoc) that lists all recodes by section. If you want to see whether there are comparable recodes between males and females or across NSFG data years, **Appendix 4** contains 3 crosswalks:

- recodes for males and females in 2011-2013 NSFG (arranged by topics)
- recodes for females across 2002, 2006-2010, and 2011-2013 NSFG (arranged by section of questionnaire)
- recodes for males across 2002, 2006-2010, and 2011-2013 NSFG (arranged by section of questionnaire)

The frequency of missing values for the recoded variables in 2011-2013 is quite low, just as it was in the 2006-2010 NSFG. Cases that have missing data on a recode (i.e., their values could not be constructed from the source variables referenced in the recode specifications) have been imputed.

Most missing recode values were assigned using **regression** imputation software in which multiple regression is used to predict a value for the case using other variables in the data

set as predictors. Regression imputation follows the same logical constraints that are built into the original recode specifications. To the extent possible, imputed values are checked to ensure that the imputed values are within acceptable ranges, and are consistent with other recodes and other data reported by the respondent.

Some cases for some recodes were imputed using **logical** imputation, which involves having a subject-matter expert at NCHS examine variables related to the variable in question, and assign a value that is consistent with those other variables and is an educated guess of the true value when there is any ambiguity. Logical imputation is also assigned as the flag value when a recode variable's value can be defined based on the imputed values of source recodes.

Imputation flag variables were created for every recode, allowing users to determine whether the value for each case is based on reported data, or imputed data. They also indicate which kind of imputation was used. Each imputation flag has the following potential values:

- 0=Questionnaire data (no imputation)
- 1=Multiple regression imputation (used most often)
- 2=Logical imputation

A value of 0 on the imputation flag means that imputation was not necessary; the reported data were sufficient to determine an appropriate value on the recode. All values other than 0 indicate that the case was imputed for this recode. Further details on the imputation process are forthcoming at the time of this writing, but since the process used for the 2011-2013 release is similar to that used in the prior release, the below report (for the 2006-2008 data) can be consulted for details:

JM Lepkowski et al. Continuous National Survey of Family Growth: Sample Design, Sampling Weights, Imputation, and Variance Estimation, 2006-2008. *Vital and Health Statistics*, Series 2(150). June 2010. Hyattsville, MD: National Center for Health Statistics. Available at: http://www.cdc.gov/nchs/data/series/sr_02/sr02_150.pdf

As noted above, all recodes have been checked thoroughly against related data items and edited if necessary for consistency. Except when it was obviously incorrect, and involved critical or commonly used variable(s), actual reported information was never replaced by an imputed value. **We recommend that analysts use all, including the imputed, cases.** Using weights and imputed cases will enable the analyst to replicate results that appear in NCHS reports. The impact of imputation on analyses can be examined by using the imputation flags.

Finding recodes in the data file and codebook: As shown in **Appendix 1** (File Indexes), the recodes and their imputation flags are clustered together near the end of each of the 3 data files. Recodes can be distinguished in the codebook documentation by the “Variable type” displayed on the codebook page between the variable name and the variable description. The word “recode” also appears at the end of the variable’s “question text” or short description.

Protections to Minimize Risk of Disclosure for Individual-Level Data

When NCHS collected data from respondents for the NSFG, those respondents were promised in the informed consent process that the information they provided would be kept confidential. NCHS is legally required to keep that promise.

Prior to their release, the proposed NSFG public use files were reviewed by the NCHS Disclosure Review Board (DRB) and the NCHS Confidentiality Officer. The NSFG staff included a number of modifications to the data files to prevent disclosure of the identities of the respondents, and at the same time preserve the analytical value of the data. In response to the DRB's review, the NSFG staff and contractor made some further changes to minimize the risk of disclosure, including the suppression or collapsing of additional variables that could be used to identify very small groups. In brief:

- All directly identifying information (such as names and addresses) has been eliminated from the public use files.
- The only geographic variable included on the public use files is a 3-category METRO recode (principal city of Metropolitan Statistical Area (MSA), other MSA, not MSA).
- All variables on the files that could potentially be used to indirectly identify individuals have been recoded or re-categorized, with particular attention to keeping categories that are substantively useful, and collapsing categories that were so small that they were of limited analytical use.

As a final step to prevent identification of individual respondents, the values of some variables have been altered for some respondents through **statistical perturbation**. That is, a few values in the data set are no longer the actual values reported by the respondents, resulting in greater uncertainty for anyone attempting to identify a particular individual they may know participated in the survey. However, these alterations, or statistical perturbations, were carefully designed to give analysts comparable statistical information as those obtained from the unaltered responses. In other words, it is extremely unlikely that either national estimates or causal models are affected by any of the alterations, except for a very slight increase in the variance of a few statistics.

For information on the issues and techniques related to disclosure limitation and confidentiality, please consult literature such as the following:

Doyle P, Lane J, Theeuwes JJM, and Zayatz LV, editors. 2001. Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies. New York: Elsevier.

Duncan GT, Stokes SL. 2004. Disclosure Risk vs. Data Utility: The R-U Confidentiality Map as Applied to Topcoding. Chance 17(3):16-20.

Federal Committee on Statistical Methodology. (May 2005). *Report on Statistical Disclosure Limitation Methodology*. (Statistical Policy Working Paper 22). Washington, DC: Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office.
<http://www.fcsm.gov/working-papers/spwp22.html>

Feinberg, SE. 2001. Statistical Perspectives on Confidentiality and Data Access in Public Health. Statistics in Medicine 20(9-10):1347-56.

Feinberg SE, McIntyre J. 2004. Data Swapping: Variations on a Theme by Dalenius and Reiss. Privacy in Statistical Databases, Proceedings: Annals of the New York Academy of Sciences 3050:14-29.

Muralidhar K, and Sarathy R. 2003. A Theoretical Basis for Perturbation Methods. Statistics and Computing 13(4):329-35.

Reiter JP. 2004. New Approaches to Data Dissemination: A Glimpse into the Future. Chance 17(3):11-15.

Trottini M, Feinberg SE. 2002. Modeling User Uncertainty for Disclosure Risk and Data Utility. International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems 10(5):511-27. October, 2002.

Zayatz, L. 2007. Disclosure Avoidance Practices and Research at the U.S. Census Bureau: An Update. Journal of Official Statistics 23(2):253-65.

The individual-level variables that could not be included on the public use NSFG files, or could not be included in their original form, are available to the research community through the NCHS Research Data Center (RDC). The full list of these variables is provided in **Appendices 7a and 7b**, and further details on the specific modifications made to these variables are provided in **Appendix 7c**. As with all data files available through the NCHS RDC, these restricted-use data are made available to researchers under special arrangements that assure confidentiality and protection of the data.

Researchers who wish to learn more about or apply for access to any of these NSFG files available through the Research Data Center – Contextual Data, Restricted-Use individual-level data, or Interviewer Observation data -- should first look at information provided on the RDC website (www.cdc.gov/rdc), and then contact either the NSFG staff at nsfg@cdc.gov or the RDC at rdca@cdc.gov.

DESCRIPTION OF CODEBOOKS

Overview

Codebooks for the NSFG provide essential information for each variable included in the public-use files. The elements of the codebook, such as variable type and universe statements are described further below. In previous NSFG public use file releases, we supplied stand-alone PDF versions of the codebooks, section by section, on the NSFG webpage in addition to the option for users to generate whatever codebook excerpts they wish via a link within the University of Michigan's online "Webdoc." For the 2011-2013 release, the stand-alone PDFs are no longer provided, but data users can still print and/or save any codebook section or series within Webdoc.

Webdoc is a Web-based tool that permits online, interactive access to the NSFG codebooks and allows easy access to all variables, quick navigation between different sections of the instrument (through hyperlinks and through menu-like lists of data files, sections and series), and searching for specific words or phrases or for specific variables. For recoded variables, direct links are provided to the recode specifications. The design and functioning of Webdoc for the 2011-2013 NSFG is essentially the same as that for Cycle 6 (2002) and 2006-2010.

To generate an excerpt for specific sections or subsections of the codebook use the hyperlink at the bottom of the screen that says "Create codebook for <Section/Subsection>." The resulting file can be saved in PDF, printed, or simply viewed on the screen. For example, at the bottom of the Webdoc page listing the variables in sub-section, "Sterilization Operations (DA)" if you click on [Create codebook for \(DA\) Sterilization Operations](#) a screen with just the variables in the DA series from Female Section D is created. You can save this codebook excerpt in PDF or other standard format, print it, or view it on the screen.

National Survey of Family Growth, 2011-2013 - Windows Internet Explorer

http://www.icpsr.umich.edu/icpsradmin/nsfg/variableGroupC

National Survey of F...

File Edit View Favorites Tools Help

Convert Select

CDC - National Center for H... NSFG - Questionnaires, Dat... NSFG 2006-2010

CDC Home | NCHS Home | NSFG Home

NSFG

2011-2013

Female Respondent File Codebook
Female Pregnancy File Codebook
Male Respondent File Codebook
Public Use Data Files, Codebooks and Documentation

National Survey of Family Growth

Search: in

All Files (search tips)

Search variable name only (search tips)

NSFG 2011-2013 :: [Female Respondent File Codebook](#) :: [SECTION D: Sterilizing Operations and Impaired Fecundity](#)

(DA) Sterilization operations

EVERTUBS -- DA-1 R Ever Had Tubes Tied/Cut/Removed
ESSURE -- DA-1b Ever had tubal sterilization procedure called ESSURE
EVERHYST -- DA-2 R Ever Had Hysterectomy
EVEROVRS -- DA-3 R Ever Had Both Ovaries Removed
EVEROTHR -- DA-4 R Ever Had Any Other Sterilizing Operation
WHOTOOPRC -- DA-5a Codes for Spontaneous Description of Other Operation
ONOTOPRC -- DA-6 R Sterile from Operation on 1 Tube/Ovary
DFNLSTRL -- DA-7 R Sterile From This Other Operation
anytubal -- R ever had a tubal sterilization (Regardless of reversal)
hyst -- R is surgically sterile at interview due to hysterectomy
ovarem -- R is surgically sterile at interview due to ovary removal
othr -- R is surgically sterile at interview due to other female operation
anyfster -- R Ever Had Sterilizing Operation (Regardless of Reversal)
ANYOPSMN -- DA-8 R's Curr H/P Ever Had Vasectomy or Other Sterilizing Operation
WHATOPSM -- DA-9 Type of Sterilizing Operation R's Current H/P Had
DFNLSTRM -- DA-10 R's Curr H/P completely sterile/impossible to become father
anymster -- R's H/P Ever Had Sterilizing Operation Regardless of Reversal
anyvas -- R's H/P ever had a vasectomy, regardless of later reversal
othrm -- R's H/P curr sterile from operation other than vasectomy

[prev](#) [next](#)

[Create codebook for \(DA\) Sterilization operations](#)

95%

Below is an example page from Webdoc displaying the detailed codebook information for the raw variable (and question) AB-1 MARSTAT. The specific elements of the codebook are described further below.

National Survey of Family Growth, 2011-2013 - Windows Internet Explorer

http://www.icpsr.umich.edu/icpsradmin/nsfg/variable/60364

National Survey of F...

File Edit View Favorites Tools Help

Convert Select

CDC - National Center for H... NSFG - Questionnaires, Dat... NSFG 2006-2010

CDC Home | NCHS Home | NSFG Home

NSFG

2011-2013

Female Respondent File Codebook
Female Pregnancy File Codebook
Male Respondent File Codebook
Public Use Data Files, Codebooks and Documentation

National Survey of Family Growth

Search: in

All Files (search tips)

Search variable name only (search tips)

NSFG 2011-2013 :: [Female Respondent File Codebook](#) :: [SECTION A: Calendar Instructions; Demographic Characteristics; Household Roster; Childhood Background](#) :: [\(AB\) Marital/cohabiting status](#)

MARSTAT (21-21)

Variable Type : raw

AB-1 : Now I'd like to ask about marital status and living together. Please look at Card 1. What is your current marital or cohabiting status?

value	label	Total
1	Married to a person of the opposite sex	1711
2	Not married but living together with a partner of the opposite sex	763
3	Widowed	21
4	Divorced or annulled	355
5	Separated, because you and your spouse are not getting along	218
6	Never been married	2532
9	Don't know	1
Total		5601

Universe : Applicable for all respondents

Notes : use recode [RMARITAL](#)

[prev](#) [next](#)

[CDC Home](#) | [NCHS Home](#) | [NSFG Home](#) | [Contact us](#)

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

95%

Elements of the Codebook

Each variable in the public use files is represented in the codebook documentation with a page or entry containing all of these elements, which are described in turn below:

- Variable name
- Variable type
- Column locations
- Question text
- Universe statement
- Response categories and unweighted frequencies
- Link to recode specifications, where applicable
- Notes, where applicable

Variable Name:

For raw and computed variables, the variable name corresponds in most cases exactly to the question or computed variable name that appears in the CAPI Reference Questionnaire (CRQ). *(For example, in the Webdoc screen capture above, the variable name and question name are the same – MARSTAT.)* Recode and intermediate variable names correspond to those found in the recode specifications (**Appendix 3**). Throughout the codebook documentation and in the recode specifications, raw and recode variables are in uppercase and computed variables are in lowercase. In some cases where questions or variables are applicable for “loops” or “arrays” (such as pregnancies, marriages, months of the year, mentions for “enter all that apply” questions, etc.), then the variable names seen in the CRQ will have numeric suffixes attached. For example, BD-8 PAYBIRTH in the CRQ for Female Section B is the question asking how the delivery costs were paid for this child’s birth, and respondents could enter all that applied. The pregnancy file includes variables for up to 4 mentions in PAYBIRTH1-4, and this information is noted in the variable labels and question text also.

Variable Type:

On the codebook page for each variable, underneath the variable name, is “Variable Type.” There are four basic variable types included in the NSFG files -- “raw,” “computed,” “recode,” or “intermediate”:

- A **raw** variable refers to a question that was asked during the interview (the majority of variables in the data files are raw). *(For example, in the Webdoc screen capture above, AB-1 MARSTAT is labeled as a raw variable.)*
- A **computed** variable is a variable that is either computed as part of the Blaise-programmed survey instrument or constructed in post-processing after data collection, from one or more raw variables. Computed variables do not have missing values imputed.
- A **recode** variable is a constructed variable created after the data is collected, from one or more raw variables, and has missing values imputed.
- An **intermediate** variable is one defined in the specifications for certain recodes. These are sometimes included in the public use data files because they can be useful for analysts. It differs from a “computed” variable because it is neither constructed during the interview, nor is it defined solely as a post-processing variable. Intermediate

variables are defined in the process of constructing a recode. They are very few in number.

Column Locations:

Next to each variable name, the codebook page gives the column locations in parentheses. *(For example, in the Webdoc screen capture above, the column location for MARSTAT is 21 on the female respondent file.)* These are primarily relevant if you wish to read in only selected variables from the ASCII data files. Program statements for SAS, Stata, and SPSS are provided on the NSFG webpage to read in each of the three NSFG data files for 2011-2013, including statements to assign appropriate variable and value labels.

Question Text:

Question text is either the actual question wording for a raw, asked question in the interview or the description of a computed, recode, or intermediate variable. The wording of the survey question is shown in the codebook for “raw” variables and is preceded by the question number. Any question wording variants are presented, sometimes in collapsed form. For computed variables (computed as part of the programmed survey instrument or in post-processing), the “question text” is the same as the variable’s short description seen in the **File Indexes (Appendix 1)**, except that we have included the “Flow Check” number from the CAPI Reference Questionnaire (CRQ) where the computed variable was defined. For example, “(Computed in Flow Check E-13b)” indicates that the variable is defined in Section E, Flow Check E-13b. CRQs can be consulted to see these flow checks and the specifications for defining each computed variable. For recodes and intermediate variables (variables not represented in the questionnaires), the question text corresponds to the variable’s short description from the Recode Specifications (see **Appendix 3**).

When variables are part of an array or loop, the question text on the codebook page indicates what is being referenced, just as the variable descriptions do in the File Indexes (**Appendix 1**). For example, the question text for male CG-5 CWPCCHSEX2 makes clear that the variable applies to the second biological child the respondent had with his current wife or cohabiting partner.

Universe Statements (“Applicable Specifications”):

In the codebook documentation, the “applicable specifications” or “universe statement” for a question indicates which respondents were asked the question or had the variable defined for them. If a question or variable was not applicable to a particular respondent, the questionnaire program skipped to the next applicable question. If a question was not skipped by any respondent or the variable was assigned a non-blank value for every case, the universe statement says, “Applicable for all respondents” or for the pregnancy file, “Applicable for all pregnancies.” *(For example, see Webdoc screen capture for AB-1 MARSTAT above.)*

Inapplicable variables are coded as blanks. Some computer programs such as SAS and Stata read a blank as a non-numeric character (a dot) or “system missing” value, but others may read it as a zero. Analysts using statistical packages other than SAS or Stata should take care to distinguish between missing values and zeroes in programs used with these data because zeroes are often valid values on NSFG variables.

For many variables in the NSFG files, an *abridged* version of the complete universe statement is provided with the core routing information. These variables have nested routing statements, and for these variables, the most proximate routing statement will be described in the universe statement. Since the universe statement contains the variable(s) that determined the routing into that question, users can trace back through the routing logic, that is, go to each preceding variable to see its routing statement and continue until the universe statement reads, “Applicable for all respondents.”

For example: the question EA-12 ECTIMESX in the female questionnaire reads, “How many different times have you used emergency contraception?” It was asked of those who had ever used emergency contraception. Thus, EA-11 MORNPIII (“whether R ever used emergency contraception”) is included in the universe statement for EA-12 ECTIMESX, and hyperlinked. Clicking on this variable takes you to the codebook page for EA-11 MORNPIII. MORNPIII was only asked of those who had ever had sex, so its universe statement contains the computed variable for ever had sex: “rhadsex,” Clicking on this takes you to the page for this computed variable which is “applicable for all respondents.”

All public use file variables referenced in the universe statements are “hyper-linked” in Webdoc (see description above) so that users can go directly to their codebook pages. The variable names will not be hyperlinked in any PDF file generated from Webdoc, but users should still find it straightforward to find the relevant codebook entries for variables referenced in the universe statement. To make this easier to locate these variables, recall that question numbers precede the names of all raw variables. Also, the names of computed variables appear in lower case, and the names of raw variables and recodes appear in upper case. If a universe statement references variables that are not included on the public-use file those variables are not hyperlinked.

In addition to consulting the universe statement or “applicable specification” in the codebook documentation, you may also wish to consult:

- The CAPI Reference Questionnaire (CRQ; see section below entitled “**Description of Questionnaires**”), which contains more detailed specifications for the questionnaire. The universe statements for the computed variables are drawn from the Flow Checks in which the variables are defined. For the raw/asked variables as well as the computed variables, the questionnaires allow you to examine the sequencing and context of your variables of interest.
- The recode specifications (**Appendix 3**), which are the source for the universe statements included in the codebook, and also for the full details on how the recode was constructed and imputed.

Response Categories and Unweighted Frequencies:

For categorical variables and several continuous variables in the NSFG, the codebook documentation lists all values with descriptive value labels and unweighted frequencies (or counts of cases). To the extent possible, the exact wording of the questionnaire response choices is shown (*for example, see Webdoc screen capture for AB-1 MARSTAT above*). Frequencies of variables that are not applicable for all respondents include the number of “inapplicable” cases. Most century month (date) and continuous variables have been collapsed *for display purposes* into more manageable groups. The original values of these variables are intact in the file unless otherwise indicated. Response categories are not displayed unless at least 1 case reported such a

response, and this also applies to “refused” (8, 98, etc.), “don’t know” (9, 99, etc.), and “not ascertained” (7, 97, etc.) responses.

Link to recode specifications: For every recode variable, there is a direct link in Webdoc to the specification for that recode in **Appendix 3**. These hyperlinks will not be active in any generated PDF files for the codebooks.

Notes (where applicable): For selected variables, the codebook page will show a Note with further information. The primary reasons for these notes are to indicate when there is a relevant recode that you should use (*see note pointing to the RMARITAL recode in the Webdoc screen capture for AB-1 MARSTAT above*), or to describe special circumstances related to the response categories (for example, see female HE-4 PLCHIV). Some notes also contain a reference to this User’s Guide, Part 2, if there is further information the user may find helpful when using the variable.

DESCRIPTION OF QUESTIONNAIRES

The NSFG webpage provides detailed outlines of the male and female questionnaires for 2011-2013, along with the full questionnaires in 2 formats:

- CAPI-Lite format
- CAPI Reference Questionnaire (CRQ) format

Both formats represent the basic content and routing of the full NSFG interviews, including the computer-assisted personal interviews (CAPI) administered by interviewers and the audio computer-assisted self-interviews (ACASI) that respondents completed on their own. However, each format of the questionnaire offers users a different level of detail on how the interview was conducted.

CAPI-Lite Format

The male and female interviews are shown in their entirety, but with abridged representations of the question wording variants and shorter descriptions of skip patterns through the interview. With this format, the emphasis is on getting a clear picture of how the questions were asked, in what order, and of which respondents, without showing every detail that was needed to program the questionnaire.

NSFG_2011-2013 FemaleCapilite.pdf
NSFG_2011-2013 MaleCapilite.pdf

CAPI Reference Questionnaire (CRQ) Format

The CRQ shows all the detailed specifications that were used to program the NSFG questionnaires in Blaise Survey Software (<http://www.blaise.com/>). While the entire female and male CAPI-Lites are contained in 2 PDF documents, the considerably longer CRQs are provided on the NSFG webpage as separate PDF files corresponding to each section of the questionnaire. For example, **NSFG_2011-2013_Female Section A_CRQ.pdf** is the CRQ for female Section A,

and **NSFG_2011-2013_Male Section F_CRQ.pdf** is the CRQ for male Section F.

- All question wording variants are shown, along with the conditions defining when each variant should be used.
- “Flow Checks” specify the precise routing through the interview based on earlier questionnaire items so that the appropriate next questions for a particular respondent appear onscreen. In addition, in some instances flow checks include the creation of new variables from one or more of the “raw” or “asked” variables. These are called “computed variables” and are described in other sections of the User’s Guide (see **Description of Codebook, “Variable Type”**). The flow check specifies in detail how these computed variables were defined. A summary list of computed variables defined in each questionnaire section can be found at the beginning of each section’s CRQ, and those that are “passed forward” to be used for routing later in the interview are listed at the end of each section’s CRQ.
- “Edit Checks,” programmed into the instrument, attempt to catch and resolve data inconsistencies during the interview, rather than requiring resolution after data collection has ended. These consistency checks are generally located in the CRQ after the questions they are intended to reconcile. They are generally scripted for ease of use, and enable the interviewer to return to specific questionnaire items and correct them, if necessary. See also the User’s Guide section on **Data Preparation for Public Use, “Logical Inconsistencies and Out-of-Range Values.”**
- Use of additional survey aids, such as Show Cards, Help Screens, and the Life History Calendar (female interview only), is noted on individual questionnaire items. For example, if a question-specific help screen was available for an item, the CRQ indicates “[HELP AVAILABLE].” If the item’s response choices were to be shown on a Show Card in the interviewer’s show card booklet, the CRQ indicates the number of the show card along with the response categories. Also shown are the onscreen instructions for interviewers that accompanied many of the questions. An example of the Life History Calendar, used only in the female interview, is shown on the next page. The version shown is for interviews conducted in 2011. Since the female interview includes a greater number of questions about dates and the relative timing of events than the male interview, the Life History Calendar was only used in the female interview. The Life History Calendar has been shown to improve recall of dates by anchoring responses to key life events for the respondent.

USER SUPPORT

Many commonly asked questions about the NSFG are addressed in the “Frequently Asked Questions” included in this User’s Guide in **Appendix 6** and throughout the main text of this User’s Guide. If, however, you have reviewed this Guide and its appendices thoroughly and you still have a question, please contact the NSFG team:

By mail: NSFG Team
 Reproductive Statistics Branch
 National Center for Health Statistics
 3311 Toledo Road
 Hyattsville, MD 20782

By email: NSFG@cdc.gov
By phone: (301) 458-4222

ACKNOWLEDGMENTS

An adequate acknowledgment of all those who made significant contributions to the design, conduct, and production of the 2011-2013 NSFG data release would require many pages. This brief acknowledgement will only name those who made the most important contributions.

The interviewing, data processing, and data file production for the 2011-2013 National Survey of Family Growth (NSFG) were conducted by the University of Michigan’s Institute for Social Research (ISR), under a contract with the National Center for Health Statistics (NCHS).

At ISR, **Mick Couper** was the Project Director. **William G. Axinn** was the Deputy Project Director. **Nicole Kirgis** and subsequently **Heidi Guyer** served as Field Director and Director of Operations, including the oversight of a group of more than 50 highly skilled interviewers and fieldwork managers. **James Wagner** was the Senior Mathematical Statistician, overseeing the sample design, weighting, imputation, and variance estimation. **Sarrah Buageila** played a key role in the development, testing, and evaluation of the survey instruments in the interviewer training and data collection phases of the project. The ISR data file and documentation production team was led by **Peter Granda**, Director of Data Processing, and included programmers **Michael Shove** and **Stephanie Windisch**. **Karl Dinkelmann** served as lead Blaise programmer for the survey instrument. The imputation programmer was **Patricia Berglund**. William Connett served as Director of Information Security and Confidentiality.

At NCHS, the NSFG team, responsible for all aspects of NSFG survey design through data file and documentation production, was comprised of **William D. Mosher** (NSFG Project Officer/Contracting Officer Representative and Team Leader during 2011-2013 interviewing, retired in August 2014), **Joyce C. Abma** (Alternate Project Officer during 2011-2013 interviewing, now Contracting Officer Representative), **Anjani Chandra** (now Team Leader), **Gladys M. Martinez**, **Jo Jones**, **Casey Copen**, **Kimberly Daniels**, **Isaedmarie Febo-**

Vazquez, Veena Billioux (until August 2012), **Jonathan Vespa** (until March 2012), and **Jill Daugherty** (since June 2013). Consultation on survey design, variance estimation and other statistical matters was provided by **Van Parsons** and **Yulei He** (since April 2013) of the NCHS Office of Research and Methodology.

The 2011-2013 NSFG was jointly planned and funded by the following agencies of the U.S. Department of Health and Human Services:

- CDC's National Center for Health Statistics (NCHS)
- Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD)
- Office of Population Affairs (OPA)
- Children's Bureau of the Administration for Children and Families (ACF)
- Office of Planning, Research, and Evaluation (OPRE/ACF)
- CDC's Division of Reproductive Health (DRH)
- CDC's Division of HIV/AIDS Prevention (DHAP)
- CDC's Division of STD Prevention (DSTDP)
- CDC's Division of Cancer Prevention and Control (DCPC)
- CDC's National Center for Birth Defects and Development Disabilities (NCBDDD)

NCHS is grateful for the support and contributions of all those noted above.

USER'S GUIDE PART 2:

TOPIC-SPECIFIC NOTES FOR ANALYSIS

ABORTION UNDER-REPORTING AMONG FEMALES IN THE NSFG

Abortions have always been under-reported in the National Survey of Family Growth (NSFG) and virtually all other demographic surveys. This has been determined by comparing NSFG weighted estimates of abortions with external data from abortion providers.

Numerator = Weighted number of abortions reported by NSFG respondents (directly to the interviewer) in a recent period such as five years prior to the start of interviewing. *(In the 2002 NSFG and later, women were re-asked in ACASI about pregnancies resulting in abortion, but the dates for these pregnancies were not asked in a uniform enough manner to include in the estimates below.)*

Denominator = Number of abortions reported in those same calendar years, based on data reported to CDC's Division of Reproductive Health or surveys of abortion providers conducted by the Guttmacher Institute.

Using this simple comparison, the estimated percentage of abortions reported by women 15-44, according to NSFG survey year, is shown below.

<u>Cycle & survey year</u>	<u>Percent reported in the NSFG</u>	
Cycle 2 (1976)	45%	
Cycle 3 (1982)	48%	
Cycle 4 (1988)	35%	
Cycle 5 (1995)	45%	
Cycle 6 (2002)	47%	(for abortions in 1997-2001)
2006-2010 NSFG	38%	(for abortions in 2001-2005)
2011-2013 NSFG	38%	(for abortions in 2006-2010)

(For further information on NSFG Cycle 6, see R.K Jones & K.Kost, "Underreporting of Induced and Spontaneous Abortion in the United States: An Analysis of the 2002 National Survey of Family Growth. *Studies in Family Planning* 38(3): 187-197. September 2007. Under-reporting based on later NSFG file releases has not yet been analyzed.) Denominators used in calculations for 2006-2010 and 2011-2013 NSFG comparisons are from R.K. Jones & J. Jerman. "Abortion Incidence and Service Availability in the United States, 2011." *Perspectives on Sexual and Reproductive Health*. 46(1):3-14. 2014.

As in previous surveys, the NSFG staff advises NSFG data users that, generally speaking, NSFG data on abortion should **not** be used for substantive research. The NSFG abortion data can be used for:

- (1) methodological studies of factors affecting abortion reporting.
- (2) studies of contraceptive efficacy, but only after the data are adjusted for the under-reporting of abortion.

The study of the determinants and consequences of abortion with these data is problematic and, in general, not advised.

BIRTH REPORTING AMONG FEMALES IN THE NSFG

The table below shows a comparison of NSFG-based estimates of births occurring in 2006-2010, relative to births registered through the vital records system in the United States for those same years. These figures are based on births reported by all female respondents in the 2011-2013 NSFG. The NSFG-based estimates are constructed from all births reported over a period of five years to increase the statistical reliability of the estimates. *(The number of births or pregnancies reported in any single calendar year by NSFG respondents would yield sampling errors too high to produce reliable estimates. Users are therefore advised to combine multiple years when analyzing these data.)* To make the NSFG estimates comparable to vital records, births are excluded if reported by non-U.S.-born NSFG respondents that occurred before they came to the U.S. to stay. Births reported in vital records are also limited to those occurring to women age 44 or younger, in order to be comparable to the NSFG age range.

For the total number of births, as well as the numbers shown by Hispanic origin and race, marital status at the time of the child's birth, mother's age at delivery, and birth order, the NSFG estimates do not differ significantly from vital records. That is, the ratios of NSFG to vital records do not differ significantly from 1, and the numbers based on vital records fall within the 95% confidence intervals for all of the NSFG estimates below.

	Number (in thousands) of births from NSFG	95% Confidence Interval Lower Limit	95% Confidence Interval Upper Limit	Vital records (in thousands) ²	Ratio of NSFG/ Vital records
All births occurring in 2006-10 ¹	21,297	18,701	23,893	20,922	1.02
Hispanic origin and race ³					
Hispanic	4,817	3,495	6,139	5,082	0.95
Not Hispanic					
White	12,295	10,593	13,996	11,239	1.09
Black or African American	2,581	1,994	3,167	3,063	0.84
Marital status at child's birth					
Married	11,794	9,924	13,664	12,518	0.94
Unmarried	9,503	7,992	11,014	8,404	1.13
Mother's age at delivery					
19 and younger	2,209	1,700	2,718	2,120	1.04
20-24	5,429	4,590	6,267	5,173	1.05
25-29	6,888	5,587	8,188	5,887	1.17
30-44	6,772	5,554	7,990	7,742	0.87
Birth order ⁴					
1 st	8,324	7,304	9,345	8,381	0.99
2 nd	7,334	6,279	8,389	6,579	1.11
3rd or higher	5,639	4,498	6,779	5,835	0.97

¹ Total includes women of other Hispanic origins and races, not shown separately.

² Vital Stats <http://www.cdc.gov/nchs/VitalStats.htm>.

³ The Hispanic origin and race variable shown here is based on the 1977 Office of Management and Budget Guidelines in order to be comparable with the Vital Statistics data.

⁴ Some births have unstated birth order in the Vital Statistics data. Those are not shown separately here.

RELIGION DATA IN THE NSFG

All NSFG respondents were asked “In what religion were you raised, if any?” and “What is your current religion, if any?” For both these questions, they were shown a card listing the 10 religious groups with the largest memberships in the United States (RELNOW and RELRSD). If their religion was not among those listed, the respondents were shown a 2nd card with 17 more religions (RELNOW1 and RELRSD1), plus an “Other-specify” option. A small number of cases provided “Other-specify” responses, and these verbatim responses (OTHRLNOW and OTHRLRSD) were reviewed to determine if any could be “back-coded” into one of the existing detailed categories or if a new category should be created. In the 2011-2013 file, there are too few cases to do analysis with the detailed religion categories.

The coding of the other-specify religion verbatim responses into existing and new categories of RELRSD/RELRSD1 (religion in which respondent was raised) and RELNOW/RELNOW1 (current religion) followed the same logic for males as for females. Although the variable names and question routing are the same in both instruments, the question numbers are different because these questions were asked in different sections – Section I for females and Section J for males. For ease of reference, the table below presents a crosswalk of the variable names with the question numbers for males and females. Besides the religious affiliation variables described above, the table also lists FUNDAM1, a choose-all-that-apply question asking whether the respondent believes certain religious labels apply to him or her.

Variable Name	Female Question Number	Male Question Number
RELRSD	IC-1	JB-1
RELRSD1	IC-2	JB-2
OTHRLRSD	IC-3	JB-3
RELNOW	IC-5	JB-5
RELNOW1	IC-6	JB-6
OTHRLNOW	IC-7	JB-7
FUNDAM1-4	IC-8	JB-8

RELCURR and RELRAISD: “Current religion” and “Religion raised”

Respondents were allowed to enter verbatim responses for the religion in which they were raised (OTHRLRSD) and the religion they practiced at the time of the interview (OTHRLNOW) if they could not find the response category for their religion on the show cards for RELRSD/RELRSD1 and RELNOW/RELNOW1.

The “Other-specify” verbatim responses for OTHRLNOW were coded into existing and new categories of RELNOW and RELNOW1. Once all verbatim responses were assigned to an existing or new category, RELNOW and RELNOW1 were used to create **RELCURR** for the public use file. The full detail of RELNOW and RELNOW1 is available through the NCHS Research Data Center (RDC).

Similarly, after “Other-specify” responses for OTHRLRSD were coded into existing and new categories of RELRSD and RELRSD1, these variables were used to create the variable **RELRAISD** for the public use file. The full detail of RELRSD and RELRSD1 is available

through the RDC.

RELCURR and RELRAISD are both variables constructed in post-processing for inclusion on the public use file, and do not have imputed values. RELCURR also serves as an intermediate variable in the construction of the RELIGION recode (see also **Appendix 3a (female)** and **3c (male) recode specifications**).

Details of the construction of RELCURR from raw variables RELNOW and RELNOW1 follow. RELRAISD was created in the same manner as RELCURR using raw variables RELRSD and RELRSD1.

- RELCURR=1 if RELNOW = None (1) or RELNOW1 = No particular faith (90)
- RELCURR=2 if RELNOW = Catholic (2)
- RELCURR=3 if RELNOW = Southern Baptist (4) or Baptist (5)
- RELCURR=4 if RELNOW = Methodist or African Methodist (6), Lutheran (7), Presbyterian (8), or Episcopal or Anglican (9)
- RELCURR=5 if RELNOW1 = Assemblies of God (12), Church of Nazarene (13), The Church of God (14), The Church of God (Cleveland, TN) (15), The Church of God in Christ (16), 7th Day Adventist (17), United Pentecostal Church (18), Pentecostal Assemblies (19), or Fundamental Protestant Bodies, Pentecostal (30)
- RELCURR=6 if RELNOW1 = Christian, another denomination not listed (21)
- RELCURR=7 if RELNOW1 = Christian, no specific denomination (22)
- RELCURR=8 if RELNOW = Jewish (3) or Church of Jesus Christ of Latter Day Saints, (LDS/Mormon) (10), or

 if RELNOW1 = Jehovah's Witness (20), Unitarian-Universalist (23), Greek Orthodox (24), Other Orthodox (25), Muslim (26), Buddhist (27), Hindu (28), Native American religions (31), Taoic religions (32), Neopagan religions (33), or Other, not shown separately (95)
- RELCURR=9 if RELNOW = Refused (98)
- RELCURR=10 if RELNOW = Don't know (99)

Code categories:

- 1 = No religion
- 2 = Catholic
- 3 = Baptist/Southern Baptist
- 4 = Methodist, Lutheran, Presbyterian, Episcopal
- 5 = Fundamental Protestant
- 6 = Other Protestant denomination
- 7 = Protestant - No specific denomination

8 = Other religion
9 = Refused
10 = Don't know

RELTRAD: “Current religious affiliation by Protestant denomination”

(Please see above table for question numbers.)

The variable RELTRAD categorizes Protestant respondents into three groups depending on the respondent's Protestant denomination (RELNOW and RELNOW1), his or her fundamentalist ideology (FUNDAM1), and race (recode RACE). RELTRAD was constructed in post-processing and missing values were not imputed. (In specifications below, “ne” means “not equal to.”) We have included RELTRAD on the 2011-2013 NSFG public use file because many scholars have found it a useful categorization of religious affiliation, “grounded in the historical development of American religious traditions” (*Steensland B, Park JZ, Regnerus MD, Robinson LD, Wilcox WB, Woodberry RD. The Measure of American Religion: Toward Improving the State of the Art. Social Forces 79(1):291-318. 2000).*

- RELTRAD=1 if RELNOW = (Southern Baptist (4) or Baptist (5)) **and** RACE ne black (1), or
- if RELNOW1 = Assemblies of God (12), Church of Nazarene (13), The Church of God (14), The Church of God (Cleveland, TN) (15), 7th Day Adventist (17), United Pentecostal Church (18), Pentecostal Assemblies (19), or Fundamental Protestant Bodies, Pentecostal (30), or
- if RELNOW1 = (Christian, another denomination not listed (21) or Christian, no specific denomination (22)), **and** RACE ne black (ne 1), **and** FUNDAM1 = a born again Christian, a charismatic, an evangelical, or a fundamentalist (1, 2, 3, 4)
- RELTRAD=2 if RELNOW = Methodist or African Methodist (6) **and** RACE ne black (ne 1), or
- if RELNOW = Lutheran (7), Presbyterian (8), Episcopal or Anglican (9), or
- RELNOW1 = (Christian, another denomination not listed (21) or Christian, no specific denomination (22)), **and** RACE ne black (ne 1), **and** FUNDAM1 = none of the above, dk, rf (5, 8, 9),
- RELTRAD=3 if RELNOW = (Southern Baptist (4), Baptist (5), or Methodist or African Methodist (6)) **and** RACE = black (1), or
- if RELNOW1 = The Church of God in Christ (16), or
- RELNOW1 = (Christian, another denomination not listed (21) or Christian, no specific denomination (22)) **and** RACE = black (1),
- RELTRAD=4 if RELNOW = Catholic (2)

- RELTRAD=5 if RELNOW = Jewish (3) or Church of Jesus Christ of Latter Day Saints, (LDS/Mormon) (10), or

if RELNOW1 = Jehovah's Witness (20); Unitarian-Universalist (23); Greek Orthodox (24); Other Orthodox (25); Muslim (26); Buddhist (27); Hindu (28); Native American religions (31); Taoic religions (32); Neopagan religions (33); or Other,-not shown separately (95)
- RELTRAD=6 if RELNOW = None (1) or RELNOW1 = No particular faith (90)
- RELTRAD=8 if RELNOW = Refused (98)
- RELTRAD=9 if RELNOW = Don't know (99)

Code categories:

- 1 = Evangelical Protestant
- 2 = Mainline Protestant
- 3 = Black Protestant
- 4 = Catholic
- 5 = Other religion
- 6 = No religious affiliation
- 8 = Refused
- 9 = Don't know

FEMALE RESPONDENT FILE NOTES

Section C: Incorrect routing on Flow Check C-76 for P1YHSAGE-P1YSHAGE3 (CI-10):

Respondents < 18 years whose most recent/last partner in the past 12 months was not a current partner were skipped erroneously past a question about the partner's age at first intercourse. Approximately 34 cases for P1YHSAGE, 33 cases for P1YHSAGE2 and 14 cases for P1YSHAGE3 were set to "97" ("not ascertained") in the 2011-2013 file. This routing error has been fixed within the survey instrument, so it will be correct in later data releases after 2011-2013 NSFG.

Section E: EA series, Open-ended questions for reasons for discontinuing specific contraceptives

The following variables are a result of coding responses to open-ended questions (3 questions for each of 4 methods).

STOPPILL1-STOPPILL6
STOPCOND1-STOPCOND2
STOPDEPO1-STOPDEPO5

STOPIUD1-STOPIUD5

The open-ended follow-up questions were triggered based on certain responses to the following variables: EA-18 REASPILL, EA-19 REASCOND, EA-20 REASDEPO, EA-21a REASIUD. In cases where the verbatim response matched an already-existing category on one of these closed-ended questions, they were coded as that category on that variable. Verbatim responses that did not appear in sufficient numbers were collapsed into “Other side effects” and “Other – (Too few cases, not classifiable elsewhere).”

Section E: Other-specify responses to EA-14 OTHRMETH (EA-15 SP_ OTHRMETH)

EA_15 SP_ OTHRMETH is the question capturing open-ended responses from respondents who chose “other method” ever used on EA-14 OTHRMETH. If the response was not among the methods already asked about in the EA series, it can be found in a new variable or variables (since EA-14 is ‘Enter all that apply’) created for this purpose: “NEWMETH.” If a response indicated a method that was already asked about it was coded as “5” on NEWMETH: “Response was coded as a method in EA-1 through EA-14.” Responses that appeared too infrequently to code separately are coded: “4” on NEWMETH: “Other method, not shown separately.”

Section E: ED Series, Contraceptive Method History

MC1MONS1, MC1MONS2, MC1MONS3

Respondents who start the method calendar in the earliest “slot” -- the month equivalent to January 3 years before the interview (cmjan3yr) -- and report using a method in that month are asked how long before that date they had been using that method. Before February 2012, respondents were asked for the number of months they had been using the method up to the date of cmjan3yr. Starting in February 2012, respondents could report either the number of months they had been using the method or the date they started using that method. There were some cases (among interviews before March 2012), where the respondent did not report the number of months, but told the interviewer the date, which-the interviewer entered as a comment. The dates entered in these comments were used to assign a number of months for the applicable MC1MONS1, MC1MONS2, or MC1MONS3 variable (number of months was calculated by: cmjan3yr minus the century month of the date in the comment). The computed variable cmdatbegin was also assigned accordingly in these cases and reflects the total universe for which it is applicable (respondents who used a method in the first slot of the method calendar, or cmjan3yr).

Monthly method history variables, METHX1-METHX192

Similar to past cycles, the ED series captures contraceptive method(s) the respondent used each month, for the time period from January, three years prior to the interview date, through interview date. For example, for interviews done in September 2011, the series collects method use information from January 2008 through September 2011. Variables for up to 4 methods for each month are saved on the data file. Therefore, variables containing these contraceptive methods range from “METHX1 through METHX192” (4 variables/methods per month, for up to 4 years, or 48 months.). With continuous interviewing, the number of calendar

years spanned by the method calendar continues to increase. The output from this series was designed to remain the same number of month/year “cells,” despite this accumulation of actual calendar years. Each respondent’s particular window of 3+ years depends on the date they were interviewed. The following table displays the correspondence between calendar month and year and the ED contraceptive variable names and how this depends on date of interview, or cmintvw.

For convenience, two sets of frequently-used computed variables based on the method calendar are included in the female respondent data file.

currmeth1 - currmeth4: the method(s) the respondent used in the month of interview.

lstmonmeth1 - lstmonmeth4: the method(s) the respondent used in the month before interview.

If interview year is **2011** (cmintvw=**1341 (September)** through **1344**), the variable names correspond to months/years as below:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2008	METHX1-4	METHX5-8	METHX9-12	METHX13-16	METHX17-20	METHX21-24	METHX25-28	METHX29-32	METHX33-36	METHX37-40	METHX41-44	METHX45-48
2009	METHX49-52	METHX53-56	METHX57-60	METHX61-64	METHX65-68	METHX69-72	METHX73-76	METHX77-80	METHX81-84	METHX85-88	METHX89-92	METHX93-96
2010	METHX97-100	METHX101-104	METHX105-108	METHX109-112	METHX113-116	METHX117-120	METHX121-124	METHX125-128	METHX129-132	METHX133-136	METHX137-140	METHX141-144
2011	METHX145-148	METHX149-152	METHX153-156	METHX157-160	METHX161-164	METHX165-168	METHX169-172	METHX173-176	METHX177-180	METHX181-184	METHX185-188	METHX189-192

If interview year is **2012** (cmintvw=**1345** through **1356**), the variable names correspond to months/years as below:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2009	METHX1-4	METHX5-8	METHX9-12	METHX13-16	METHX17-20	METHX21-24	METHX25-28	METHX29-32	METHX33-36	METHX37-40	METHX41-44	METHX45-48
2010	METHX49-52	METHX53-56	METHX57-60	METHX61-64	METHX65-68	METHX69-72	METHX73-76	METHX77-80	METHX81-84	METHX85-88	METHX89-92	METHX93-96
2011	METHX97-100	METHX101-104	METHX105-108	METHX109-112	METHX113-116	METHX117-120	METHX121-124	METHX125-128	METHX129-132	METHX133-136	METHX137-140	METHX141-144
2012	METHX145-148	METHX149-152	METHX153-156	METHX157-160	METHX161-164	METHX165-168	METHX169-172	METHX173-176	METHX177-180	METHX181-184	METHX185-188	METHX189-192

If interview year is **2013** (cmintvw=**1357** through **1365 (September)**), the variable names correspond to months/years as below:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2010	METHX1-4	METHX5-8	METHX9-12	METHX13-16	METHX17-20	METHX21-24	METHX25-28	METHX29-32	METHX33-36	METHX37-40	METHX41-44	METHX45-48
2011	METHX49-52	METHX53-56	METHX57-60	METHX61-64	METHX65-68	METHX69-72	METHX73-76	METHX77-80	METHX81-84	METHX85-88	METHX89-92	METHX93-96
2012	METHX97-100	METHX101-104	METHX105-108	METHX109-112	METHX113-116	METHX117-120	METHX121-124	METHX125-128	METHX129-132	METHX133-136	METHX137-140	METHX141-144
2013	METHX145-148	METHX149-152	METHX153-156	METHX157-160	METHX161-164	METHX165-168	METHX169-172	METHX173-176	METHX177-180	METHX181-184	METHX185-188	METHX189-192

Section F: Mapping of family planning variables

In female Section F we ask about a set of 14 different possible services women could have received in the past 12 months. For each service received we gather additional information specific to that service: type of provider, whether that provider is the regular source of care, source of payment, etc. This table shows how the different variables are mapped by the type of service received.

Mapping of family planning service variables in 2011-2013 NSFG data release						
Service received in last 12 months	Question number	Service variable name	Recodes	Provider	Regular source of care	Source of payment
Method Birth Control/Prescription	FA-1b	BTHCON12	FPTITBC	BC12PLCX2	REGCAR12_F_02	BC12PAYX7-12
Checkup for Birth Control	FA-1c	MEDTST12	FPTITCHK	BC12PLCX 3	REGCAR12_F_03	BC12PAYX13-18
Counseling About Birth Control	FA-1d	BCCNS12	FPTITCBC	BC12PLCX 4	REGCAR12_F_04	BC12PAYX19-24
Sterilizing operation	FA-1e	STEROP12	FPTITSTE	BC12PLCX 5	REGCAR2_F_05	BC12PAYX25-30
Counseling re Getting Sterilized	FA-1f	STCNS12	FPTITCST	BC12PLCX 6	REGCAR12_F_06	BC12PAYX31-36
Emergency Contraception/Prescription	FA-1g	EMCON12	FPTITEC	BC12PLCX 7	REGCAR12_F_07	BC12PAYX37-42
Counseling re Emergency Contraception	FA-1h	ECCNS12	FPTITCEC	BC12PLCX 8	REGCAR12_F_08	BC12PAYX43-48
Pregnancy Test	FA-3a	PRGTST12	FPTITPRE	BC12PLCX 9	REGCAR12_F_09	BC12PAYX49-54
Abortion	FA-3b	ABORT12	FPTITABO	BC12PLCX 10	REGCAR12_F_10	BC12PAYX55-60
Pap Smear	FA-3c	PAP12	FPTITPAP	BC12PLCX 11	REGCAR22_F_11	BC12PAYX61-66
Pelvic Exam	FA-3d	PELVIC12	FPTITPEL	BC12PLCX 12	REGCAR2_F_12	BC12PAYX67-72
Prenatal Care	FA-3e	PRENAT12	FPTITPRN	BC12PLCX 13	REGCAR2_F_13	BC12PAYX73-78
Post-Pregnancy Care	FA-3f	PARTUM12	FPTITPPR	BC12PLCX 14	REGCAR2_F_14	BC12PAYX79-84
Counsel/test/treatment for STD	FA-3g	STDSVC12	FPTITSTD	BC12PLCX 15	REGCAR2_F_15	BC12PAYX85-90
All services in one visit:	FA-4	NUMBCVIS=1 all in one visit	-	BC12PLCX	REGCAR12_F_01	BC12PAYX- BC12PAYX 6

Female Recode LSEXAGE: Respondents with values of 45 years old

On the female recode LSEXAGE (respondent's age at last sexual intercourse), 4 respondents in 2011-2013 have a value of 45 years. Two of these respondents were 44 years old at the time of the screener (computed variable agescrn=44), and had their 45th birthday before the interview was conducted (computed variable age_r and recode AGER=45), so LSEXAGE is correctly equal to 45.

For the other 2 respondents, their age at last intercourse LSEXAGE was actually 44, but was computed to be 45 based on the fact that last intercourse occurred in the month of interview, which happened to be their birth month. These respondents reported in question AA-1 AGE_A that they were 44 years old, indicating that the interview occurred before their actual birthday in that month. However, the formula for computing LSEXAGE does not permit this level of precision because it is based only on century-months, not month/day/year. The formula yielded 45 years because it simply subtracts the century-month of the respondent's birth from the century-month of her last sexual intercourse and divides by 12 to obtain the age in years.

Female LSEXPAGE: Reverted from recode to "computed in post-processing" due to imputation challenges

This variable was originally intended to be a recode that would have missing values imputed. However, a routing error caused unexpected levels of missing data for respondents younger than 18 years old on a key input variable (CI-10 PIYHSAGEX) (see "Incorrect routing on Flow Check C-76 for PIYHSAGE-PIYSHAGE3 (CI-10)" in this User's Guide). It was judged too risky to impute these data with the potential to introduce a large number of unrealistic age combinations for minor respondents and their partners at last sex. Therefore the missing data were not imputed, and these cases were assigned "997" codes for "not ascertained." The following are the specifications for the construction of this variable if the users wish to create it on their own:

LSEXPAGE is blank (inapplicable) if R has never had sexual intercourse (recode HADSEX=2).

If R's most recent partner was R's first partner ever (LIFEPR=1 or MATCHFP=1 or p1yrelp=13 or (p1yrelp=7 and CG-2 SAMEMAN=1 and CA-1 TIMESMAR=1) or (p1yrelp=8 and CG-2 SAMEMAN=1 and TIMESCOH=1)):

If R had sex only once with first partner (cmlsxfp=9996) then assign partner's age at first sex with R:
LSEXPAGE = recode **FSEXPAGE**

Else, determine # of months that elapsed between first and last sex with first partner:
calculate elapsed1:
If cmlsxfp not sysmis, 9998, or 9999 and cmfstsex not sysmis, 9998, or 9999 then:
elapsed1 = cmlsxfp minus cmfstsex

Then, estimate partner's age at last sex based on elapsed1 value and partner's age at first sex with R, recode **FSEXPAGE**, and flag with leading 9:

LSEXPAGE = **FSEXPAGE** + INT(elapsed1/12)) + 900

Else if R's most recent partner was not R's first partner ever (LIFEPR>1 and p1yrelp ne 13):

[Note: for those age 18 or over, and those under 18 but partner not current, all we have is partner's age at first sex with R. Use Section C variable for partner's age at first sex, and if had sex more than once with him, add the number of months elapsed between first & last sex to this age. For those under 18 whose partner is current, all we have is partner's age at last sex with R. Use Section J variable for partner's age at last sex with R (doesn't matter whether had sex with him more than once)]

If R is 18 or older (age_r ge 18) or

R is under age 18 (age_r < 18) and partner is not current (CI-7 P1YCURRPX ne 1 or parts12=0), then use age of most recent partner at first sex with R, from Section C:

If R had sex only once with most recent partner (cmfsex=9996) then assign (valid) partner's age at first sex with R:

if 5 le CI-10 P1YHSAGEX le 95 then:

LSEXPAGE = CI-10 P1YHSAGEX

Else, determine # of months that elapsed between first and last sex with most recent partner:
calculate elapsed2:

If cmfsex not sysmis, 9998, or 9999 and cmlsex not sysmis, 9998, or 9999 then:

elapsed2 = cmlsex minus cmfsex

Then, estimate partner's age at last sex based on elapsed2 value and (valid) partner's age at first sex with R, and flag with leading 9:

if 5 le CI-10 P1YHSAGEX le 95 then:

LSEXPAGE = CI-10 P1YHSAGEX + INT(elapsed2/12)) + 900

Else if R is under age 18 (age_r < 18) and partner is current (CI-7 P1YCURRPX=1), then use (valid) age of most recent partner at last sex, from ACASI, section J:

If JF-2a CURRPAGE <= 95 then LSEXPAGE=KG-3a CURRPAGE

Else if JF-2a CURRPAGE = DK/RF then:

[If partner was 1-2 years older, add 2 years to R's age at 1st sex and flag with leading 9]

if (JF-2b RELAGE=1 and JF-2c HOWMUCH=1) then LSEXPAGE=(VRY1STAG + 2) + 900

[If partner was 3-5 years older, add 4 years to R's age at 1st sex and flag with leading 9]

if (JF-2b RELAGE=1 and JF-2c HOWMUCH=2) then LSEXPAGE=(VRY1STAG + 4) + 900

[If partner was 6-10 years older, add 8 years to R's age at 1st sex and flag with leading 9]

if (JF-2b RELAGE=1 and JF-2c HOWMUCH=3) then LSEXPAGE=(VRY1STAG + 8) + 900

[If partner was more than 10 years older, add 10 years to R's age at 1st sex and flag with leading 9]

if (JF-2b RELAGE=1 and JF-2c HOWMUCH=4) then LSEXPAGE=(VRY1STAG + 10) + 900

[If partner was 1-2 years younger, subtract 2 years from R's age at 1st sex and flag with leading 9]

if (JF-2b RELAGE=2 and JF-2c HOWMUCH=1) then LSEXPAGE=(VRY1STAG - 2) + 900

[If partner was 3-5 years younger, subtract 4 years from R's age at 1st sex and flag with leading 9]

if (JF-2b RELAGE=2 and JF-2c HOWMUCH=2) then LSEXPAGE=(VRY1STAG - 4) + 900

[If partner was 6-10 years younger, subtract 8 years from R's age at 1st sex and flag with leading 9]

if (JF-2b RELAGE=2 and JF-2c HOWMUCH=3) then LSEXPAGE=(VRY1STAG - 8) + 900

[If partner was more than 10 years younger, subtract 10 years from R's age at 1st sex and flag with leading 9]

if (JF-2b RELAGE=2 and JF-2c HOWMUCH=4) then LSEXPAGE=(VRY1STAG - 10) + 900

[If partner was the same age, then use R's age at last sex and flag with leading 9]

if JF-2b RELAGE=3 then LSEXPAGE=VRY1STAG+ 900

Notes:

- Under some conditions, computed based on imputed values of FSEXPAGE.
- “don’t know followup” questions RELAGE and HOWMUCH: in some cases they have valid data but were not used because the combinations of values didn’t meet the criteria above.

Code categories:

Blank = Inapplicable
 00-95 = Partner’s age at last sexual intercourse, reported
 900-995=Partner’s age at last sexual intercourse, estimated
 997 = Not ascertained

FEMALE PREGNANCY (INTERVAL) FILE NOTES

Section B (Pregnancy History) verification of key pregnancy data:

In the 2011-2013 NSFG, as in past NSFG surveys, all female respondents were asked questions about each of the pregnancies they have had, whether current or completed. They were asked to report these pregnancies in chronological order, starting with the first pregnancy, however in every survey period, some women did not report them in this order. Before moving on to questions that are based on pregnancies and pregnancy intervals in correct chronological order, it is critical that the pregnancy information is sorted manually (as in older Cycles before CAPI programming was introduced) or sorted by the programmed instrument (as in Cycle 5 in 1995).

In 2011-2013, a similar approach was taken as in 2006-2010, for sorting and correcting the pregnancy-specific information that women reported, and it was used to improve both the accuracy and completeness of the pregnancy data.

- After *each completed pregnancy loop*, women were asked to confirm key data about the pregnancy in the BH-6 CNFMPREG question – specifically, the pregnancy outcome, end date, and gestational length. The question wording for CNFMPREG was tailored to the type of outcome. For example, if the pregnancy ended in a single live birth, women were asked:

“This pregnancy ended in the birth of 1 baby named [*first name or initials of baby*]. This pregnancy lasted [*X*] months and [*Y*] weeks and ended in [*baby’s month/year of birth*]. Is this correct?

If the woman said “yes,” then the interviewer moved on the next pregnancy. If she said “no,” she was given a chance to correct any information on outcome, end date, and gestational length.

- After all pregnancies were looped through the pregnancy-specific questions, women were asked to verify the chronological order of all of their pregnancies (BI-2 CHKORDER). Any current pregnancy was assumed to be the most recent pregnancy. All of the information for a woman’s pregnancies are presented in a summary screen (see CAPI Reference Questionnaire for details), and she is presented with a summary statement about each pregnancy, similar to BH-6 CNFMPREG above. At this point, the respondent is given the following choices:
 - Yes, pregnancies in order/everything is correct
 - No, pregnancies out of order

She may also volunteer that something else is incorrect; for example, she may decide to add or delete a pregnancy. If she reports that her pregnancies are out of order, she is asked BI-8 FIXORDER to place them all in the correct order. If any other key information about her pregnancies is incorrect,

she is guided through specific questions to correct the information on outcome, end date, and gestational length.

In any case, whether the respondent corrects any information or simply confirms that all is correct, **new variables** are created to hold the “corrected and chronologically sorted” pregnancy history data, or “**CCSD**” as the variable labels will indicate. These CCSD variables are passed forward to all later sections of the interview and are used in subsequent routing and consistency checks. The pregnancy order variable (PREGORDR) and recode for birth order (BIRTHORD) on the public use file are both based on CCSD data.

The original variables as initially reported by the respondent *and* the CCSD computed variables, which generally end in “_s,” are included on the public use file. **For the vast majority of pregnancies reported by NSFG respondents, the CCSD variables are identical to the originally reported variables because no correction was needed.**

The table below provides a crosswalk of the original variables, the CCSD variables, and the corresponding recodes on the pregnancy or respondent file. All of the CCSD variables listed are defined in Flow Check B-42d (see CAPI Reference Questionnaire, Female Section B for details). The Flow Checks (FC) where the other computed variables are defined is listed in the table. As indicated throughout this User’s Guide, it is best to use the official recodes when available, as they have had missing values imputed in a well-documented and consistent manner and will allow you to match your figures to published statistics. See **Appendix 3a (female respondent)** or **3b (pregnancy) recode specifications** for further details.

File	Short Description	Originally Reported Variable	Corrected and Chronologically Sorted Data (CCSD)	Recode
Resp	Number of pregnancies	BB-1 NUMPREGS	npregs_s	PREGNUM
Resp	Number of liveborn babies	numbabes (FC B-1)	nbabes_s	PARITY
Preg	Outcome of pregnancy	prgoutcome (FC B-42a)	outcom_s	OUTCOME
Preg	End date of pregnancy	cmprgend (FC B-42a)	cmendp_s	DATEND
Preg	Start date of pregnancy	cmprgbeg (FC B-42a)	cmpbeg_s	DATECON
Preg	Number of babies born alive from this pregnancy	bornaliv (FC B-10)	nbrnlv_s	n/a
Preg	Gestational length in weeks	wksgest (FC B-13)	n/a	PRGLNGTH

MALE FILE NOTES

Biological children in chronological order:

In the male questionnaire, respondents were asked about their biological children in the context of questions about their relationship with the mothers of these children. Based on consultation (before men were first included in the 2002 NSFG) with experts in surveys of men, this approach was considered to yield the most accurate reporting of men's sexual, fertility, and contraception experiences. However, with this approach, questionnaire information on a biological child is located on the data files in the section where a man reported the child:

- In Section C if the child's mother is his current wife or cohabiting partner
- In Section D if the child's mother was his last sexual partner ever, or one of his 3 most recent partners in the last 12 months
- In Section E if the child's mother was a former wife or his 1st cohabiting partner (and not a recent sexual partner)
- In Section F if the child's mother was any other sexual partner not discussed in Sections C-E

In each of the “woman-specific” loops in Sections C-E and in Section F, space was allowed for up to 10 biological children to be reported – for example, up to 10 children could be reported with the current wife or cohabiting partner in Section C, up to 10 with *each* former wife reported in Section E. In the 2011-2013 NSFG, there were never more than 9 children reported across all of these woman-specific loops in Sections C-E or in Section F. See exact numbers below for the reporting of biological children fathered across Section C-F among male respondents in the 2011-2013 NSFG:

- **Section C** –no respondent reported more than 9 children with his current wife or cohabiting partner
- **Section D** –
 - No more than 4 children reported with “P1” most recent partner
 - No more than 5 children reported with “P2” 2nd-most-recent partner in past year
 - No more than 2 children reported with “P3” 3rd-most-recent partner in past year
- **Section E** –no respondent reported more than 4 former wives were reported in the 2011-2013 NSFG and:
 - No more than 5 children reported with “W1” 1st former wife
 - No more than 3 children reported with “W2” 2nd former wife
 - No more than 3 children reported with “W3” 3rd former wife
 - No children reported with “W4” 4th former wife
 - No more than 6 children reported with “1st Cohab Partner”
- **Section F** –no respondent reported more than 8 biological children in this section

To assist users who wish to analyze information on men's biological children based on chronological order of birth, the male data file includes selected variables derived from the section-specific biological child variables in Sections C-F. These chronologically ordered variables are located on the male file at the end of Section F questionnaire items. Each of the characteristics listed below, with the exception of BIOLIVNGnn, are arranged as arrays of 9 variables each because no male respondent reported fathering more than 9 biological children in the 2011-2013 NSFG.

The chronologically-based BIO child variables included on male file are:

- BIODOBn: Century month of nth child's birth
- BIOSEXn: Sex of nth child

- BIOAGEn: Age of nth child
- BIOAGEGPn: Age group of nth child
- BIOHHn: Whether nth child lives in same household with R
- BIOMOMn: R's relationship to biological mother of this nth child (*this array is defined based on what section of the questionnaire the child was reported in*)
- BIOMARn: Whether R was married to nth child's mother at time of child's birth
- BIOCOHBn: Whether R was living with nth child's mother at time of child's birth (includes cohabiting or married)
- BIOLRNPGn: When R learned of the pregnancy (before or after child was born)
- BIOLIVNGnn: Where child is living now (4 mentions for each of up to 9 children)
- BIOCHSIGn: If nth child born outside marriage: Whether R signed application for birth certificate or other paternity document
- BIOCHCRTn: If nth child born outside marriage: Whether R went to court to establish paternity
- BIOCHGENn: If nth child born outside marriage: Whether R established paternity by blood or genetic test
- BIOLVEVRn: Whether R ever lived with nth child (if not living with child now)
- BIOHWFARn: How far away nth child lives (in miles) from respondent
- BIOWANTn: For nth child: Wantedness of the pregnancy by R right before it began
- BIOHSOONn: For nth child: Timing of the pregnancy right before it began
- BIOHPYPGn: For nth child: R's happiness about the pregnancy when he learned about it

The table on the next page illustrates how these chronologically arranged variables are derived from questionnaire items in Sections C-F. Please consult the codebook to see further details on universe statements and response categories for these variables or the male questionnaire CRQs for the precise wording and routing for each item in Sections C-F.

In summary:

These chronologically ordered variables are essentially identical in content to the source variables in Sections C-F. They are provided for the user's convenience and hold no different values than the originally collected data in Sections C-F; they only arrange the information in a different (chronological) manner. Users whose primary goal is to examine data on men's biological children in order of their birth may prefer to use these chronologically arranged variables. If, however, their primary goal is to examine men's fertility in the context of their relationships with their children's mothers and they wish to incorporate other variables describing those relationships, it may be easier or more appropriate to use the source variables in Sections C-F.

Chronologically arranged BIO variable	Brief Description	Section C source	Section D source	Section E source	Section F source
BIODOB[x]	Century Month date of birth	CG-6 CWPCHDOB_M/Y yields cmchdob[x]	DH-6 PXCXBORN_M/Y yields cmchdob[x]	ED-6 FWPCHDOB_M/Y yields cmchdob[x]	FA-9m/y OBCDOB_M/Y yields cmchdob[x]
BIOSEX[x]	Sex of child	CG-5 CWPCHSEX	DH-5 PXCXSEX	ED-5 FWPCHSEX	FA-8 OBCSEXX
BIOAGE[x]	Age of child in years	based on cmchdob[x]	based on cmchdob[x]	based on cmchdob[x]	based on cmchdob[x]
BIOAGEGP[x]	Age group of child	CG-12 CWPCHAGE	DH-12 PXCXAGE	ED-12 FWPCHAGE	FA-15 OBCAGE
BIOHH[x]	Is child in R's household	based on CG-11 CWPCHLIV	based on DH-11 PXCXLIV	based on ED-11 FWPCHLIV	based on FA-14 OBCLIV
BIOMOM[x]	Relationship of child's mother to R (coded based on which section child was reported in)	set to 1 if current wife; set to 2 if current cohab	set to 3 if recent or last partner & also a former wife/cohab partner; set to 4 if any other recent or last partner	set to 5 if former wife; set to 6 if 1st cohab (but not a former wife)	set to 7 if child was reported in this section or not otherwise coded
BIOMAR[x]	Was R married to mother at time of child's birth	CG-8 CWPCHMAR	DH-8 PXCXMARB	ED-8 FWCHMARB	Inapp (all nonmar)
BIOCOHB[x]	Was R living with mother at time of child's birth (married or cohabiting)	CG-8 CWPCHMAR & CG-9 CWPCHRES	DH-8 PXCXMARB & DH-9 PXCXRES	ED-8 FWCHMARB & ED-9 FWPCHRES	FA-12 OBCMLIV
BIOLRNPG[x]	When R learned of the pregnancy	CG-10 CWPCHLRN	DH-10 PXCXKNOW	ED-10 FWPCHLRN	FA-13 OBCKNOWX
BIOLIVNG[x]	Where child living now (up to 4 mentions for each child)	CG-11 CWPCHLIV	DH-11 PXCXLIV	ED-11 FWPCHLIV	FA-14 OBCLIV
BIOCHSIG[x]	Whether established paternity by signing birth certificate application	CG-13a CWPCHSIG	DH-13a PXCXSIG	ED-13a FWPCHSIG	FA-16A OBCCHSIG
BIOHCRT[x]	Whether established paternity by going to court	CG-13b CWPCHCRT	DH-13b PXCXCRT	ED-13b FWPCHCRT	FA-16b OBCCHCRT
BIOCHGEN[x]	Whether established paternity by blood or genetic test	CG-14 CWPCHGEN	DH-14 PXCXGEN	ED-14 FWPCHGEN	FA-17 OBCCHGEN
BIOLVEVR[x]	Has R ever lived with child (if not living with child now)	CG-15 CWPCHVEVR	DH-15 PXCXEVR	ED-15 FWPCHVEVR	FA-18 OBCEVER
BIOHWFAR[x]	How far away does child live (miles)	CG-16 CWPCHFAR	DH-16 PXCXFAR	ED-16 FWPCHFAR	FA-19 OBCFAR
BIOWANT[x]	Wantedness of this pregnancy	CG-17 CWPCHWNT	DH-17 PXWANT	ED-17 FWPRWANT	FA-20 OBCRWANX
BIOHSON[x]	Timing of this pregnancy	CG-18 CWPCHSON	DH-18 PXSOON	ED-18 FWPSON	FA-21 OBCSOONX
BIOHPYPG[x]	R's happiness about this pregnancy	CG-19 CWPCHHPY	DH-19 PXHPYPG	ED-19 FWPHPYPG	FA-22 OBCHPYX

Nonbiological children cared for by men (Sections CI, CJ, DJ, DK, EE, EF, and FB)

In the 2011-2013 NSFG, men were asked for the number of nonbiological children he had cared for 1) within each of his marriages or cohabitations and 2) outside of any relationship with a female partner. For every nonbiological child who lived with the man, either those that his wife or partner had prior to their relationship or those not related to either the man or his wife/partner for whom they cared, information about each child's sex, current residence, and age was obtained. In addition, men were asked whether they had become the child's legal guardian or had adopted or were trying to adopt the child.

Reporting on contraceptive use by female partners (Sections C and D)

In the 2006-2010 questionnaire, questions asking whether the respondent's wife/cohabiting partner/partner used a method of contraception at last sex, and if so, which method, included a "Don't Know" category among the explicit response options presented on the screen. This reflected the fact that the respondent is reporting on another person's behavior and may not have the information being asked for, making "don't know" a legitimate response. To facilitate the interviewing protocol for the 2011-2013 NSFG, this "don't know" category was removed, so interviewers entered only the standard "Ctrl-D" for a don't know response. This affects CE-7 CWPLUSE2, CE-8 CWPLMET2, DD-7 PXLPUSE, and DD-8 PXLPMETH.

Recode LSEXRAGE (based on data from sections B-D): Respondents with values of 45 years old

On the male recode LSEXRAGE (respondent's age at last sexual intercourse), 6 respondents in 2011-2013 have a value of 45 years. Three of these respondents were 44 years old at the time of the screener (computed variable `agescrn=44`), and had their 45th birthday before the interview was conducted (computed variable `age_r` and recode `AGER=45`), so LSEXRAGE is correctly equal to 45.

For the other 3 respondents, their age at last intercourse LSEXRAGE was actually 44, but was computed to be 45 based on the fact that last intercourse occurred in the month of interview, which happened to be their birth month. These respondents reported in question AA-1 AGE_A that they were 44 years old, indicating that the interview occurred before their actual birthday in that month. However, the formula for computing LSEXRAGE does not permit this level of precision because it is based only on century-months, not month/day/year. The formula yielded 45 years because it simply subtracts the century-month of the respondent's birth from the century-month of her last sexual intercourse and divides by 12 to obtain the age in years.